



Nils Mesterton

Paikkatietojen automaattinen laadunarviointi avoimen lähdekoodin ohjelmistoilla

Diplomityö, joka on jätetty opinnäytteenä tarkastettavaksi diplomi-insinöörin tutkintoa varten.

Espoo, marraskuu 2015

Valvoja: Professori Henrik Haggrén

Ohjaaja: TkT Antti Jakobsson

Tekijä Nils Mesterton

Työn nimi Paikkatietojen automaattinen laadunarviointi avoimen lähdekoodin ohjelmistoilla

Koulutusohjelma Geomatiikka

Pääaine Fotogrammetria ja kaukokartoitus

Koodi M3006

Työn valvoja Professori Henrik Haggrén

Työn ohjaaja(t) TkT Antti Jakobsson

Päivämäärä 30.11.2015

Sivumäärä 57 + 17

Kieli Suomi

Tiivistelmä

Eurooppalaiset maanmittauslaitokset ovat luomassa yhteistyömaiden kansallisista aineistoista koostuvaa, maiden rajojen yli harmonisoitua viitekehystä European Location Framework (ELF) -projektissa. Kolmivuotiseen projektiin osallistuu 30 organisaatiota eri puolilta Eurooppaa. Projektin tavoitteena on tuoda INSPIRE-yhteensopivia aineistoja paremmin saataville. Aineistojen laadunarviointi on olennainen osa harmonisointia, sillä aineistojen laatutiedot ovat käyttäjälle tärkeitä arvioidessaan tietoaineistojen soveltuvuutta käyttötarkoitukseensa. ELF-projektissa automaattisen laadunarvioinnin avulla varmistetaan, että yhteiseen tietoaineistoon liitetään vain määrittelyiden mukaista aineistoa. Automaattinen laadunarviointi mahdollistaa nopean koko tietokannan laajuisen eheystarkastelun, jota voidaan tehdä kaikissa tietoaineiston elinkaaren vaiheissa.

Työn tavoitteena oli tutkia, voidaanko paikkatietojen automaattinen laadunarviointi toteuttaa ELF-projektin laatuvaatimusmäärittelyiden pohjalta avoimen lähdekoodin ohjelmistoilla. Määrittelyiden mukaiset laatutestit pyrittiin toteuttamaan PostGIS-tietokantaohjelmistolla. Tietokannan hallintaan sekä aineiston käsittelyyn ja tarkasteluun käytettiin Quantum GIS -paikkatieto-ohjelmistoa.

Työn tulokset osoittavat, että PostGIS tarjoaa kaikki laadunarvioinnissa tarvittavat analyysiominaisuudet. Tuloksissa epävarmuus kohdistuu ohjelmiston suorituskykyyn, sillä joidenkin laatutestien suoritusajat olivat pitkiä. Parempi luotettavuus tuloksille saataisiin vertaamalla työssä esitettyä ratkaisua johonkin muuhun ohjelmistoon.

Jatkokehitystä varten testeille esitettiin optimointimenetelmiä. Todettiin myös, että työn perusteella voidaan toteuttaa kaikki määrittelyiden mukaiset testit kattava avoimen lähdekoodin ohjelmointikirjasto tai laadunarviointiohjelmisto.

Avainsanat Laadunarviointi, paikkatieto, PostGIS, PostgreSQL, avoin lähdekoodi, ISO, OGC, ELF, ESDIN, INSPIRE

Author Nils Mesterton		
Title of thesis Automatic spatial data quality validation using open source software		
Degree programme Degree Programme in Geomatics		
Major Photogrammetry and Remote Sensing		Code M3006
Thesis supervisor Professor Henrik Haggrén		
Thesis advisor(s) D.Sc. (Tech.) Antti Jakobsson		
Date 30.11.2015	Number of pages 57 + 17	Language Finnish

Abstract

European National Mapping and Cadastral Authorities (NMCAs) are in the process of creating a harmonized pan-European spatial data framework in the European Location Framework (ELF) project. ELF project consortium consists of 30 organizations, and the project is set to be completed by 2016. The main goal of the project is to enable easier access to harmonized INSPIRE data in cross-border and pan-European use cases. Data quality validation is an important part of the harmonization process, making it easier for a user to determine whether or not a dataset can be utilized for a specific purpose. Automatic data quality validation ensures that a dataset complies with data requirements before it's added to the ELF. Automatic data quality validation is a relatively quick method for assessing spatial data consistency even throughout an entire database. This adaptive method can be applied on every phase of the spatial data life cycle.

The objective of this thesis was to evaluate if open source tools could be used for enabling automatic data quality validation on ELF-compatible spatial data. PostGIS database software was used to implement quality tests based on ELF data quality requirements. Quantum GIS software was used for managing the database and for viewing, processing and constructing test data.

Results indicate that PostGIS offers all analysis capabilities required for validating consistency of spatial data. Uncertainties lie in the performance aspect of the test environment because some of the tests indicated lengthy processing times. Reliability of the research could be improved by comparing the presented solution to an existing software product.

Optimization methods are presented for further development. It is also suggested that the presented method would be suitable for developing an open source software library or a quality validation tool.

Keywords Data quality validation, spatial data, PostGIS, PostgreSQL, open source, ISO, OGC, ELF, ESDIN, INSPIRE

Alkusanat

Diplomityöpaikkaa hakiessani halusin yhdistää työn aiheessa kiinnostukseni ohjelmisto- ja maanmittausaloja kohtaan. Kuin kohtalon saattelemana päädyin ELF-projektiin selvittämään, voidaanko Suomen kansallisten paikkatietoaineistojen laadunarviointi suorittaa avoimen lähdekoodin ohjelmistoilla. Diplomityössäni pääsin lopulta uppoutumaan opintojen aikaisten kiinnostusteni lisäksi omassa arvoasteikossani hyvin tärkeään asiaan, eli laatuun. Työ osoittautui suureksi urakaksi etenkin loppuvaiheessa, mutta prosessin myötä sain oppia valtavasti uusia ja jännittäviä asioita. Varsinkin hauskaa laatutestien kehittelyä tulen muistamaan lämmöllä vielä pitkään.

Kiitokset työn ohjaamisesta kuuluvat Antti Jakobssonille ja työn valvomisesta Henrik Haggrénille. Kiitos Juha Vilhomaalle, joka alkujaan ohjasi viestini eteenpäin MML:n sisällä. Kiitän myös kaikkia muita Paikkatietoinfrastruktuurin palvelut -osaston ja Oskari-tiimin henkilöitä, jotka auttoivat työtä edistymään, ja jotka olivat syyllisiä kivaan työilmapiiriin. Kiitokset kuuluvat tietysti myös elämäni vakiintuneille kulmakiville, eli kaikille, jotka enemmän tai vähemmän pakosta ovat vuosien varrella lähimaastossani joutuneet pyörimään. Tähän joukkoon kuuluvat ainakin lähin perhe ja ehkä muutkin sukulaiset, mahdollisesti myös ystävät sekä erinäinen joukko muitakin. Kaikille ehdottomasti kiitos, jotka tuntevat sen ansainneensa. Erikseen haluan kiittää kaikesta vielä Äitiä.

Seuraava haasteeni on navigoida reippaaseen uran alkuun, mutta uskoisin maanmittarin koulutuksen saaneella olevan hyvät edellytykset määrittämään paikkansa valmistuttuaan.

Kirkkonummella 30.11.2015

Nils Mesterton

Sisällysluettelo

1	Johdanto	1
1.1	Työn tausta	1
1.2	Tutkimuksen tavoitteet ja rajaukset	2
1.3	Työn rakenne	2
2	Paikkatiedot ja niiden laatu	4
2.1	Laatu käsitteenä	4
2.2	Paikkatietojen olemus ja virheet	4
2.3	Koordinaatit paikkatietojärjestelmissä	5
2.4	Standardit ja suositukset	6
2.5	Laatutekijät	6
2.6	Laatumittarit ja laadunarviointimenetelmät	8
2.7	Metalaatu	10
2.8	Laatusäännöt	10
2.9	Työvälineitä laadun arviointiin	12
3	Spatiaaliset relaatiot	14
3.1	Tausta	14
3.2	Paikkatietokohteiden geometriset kuvaukset	14
3.3	Spatiaalisten relaatioiden esitystavat	15
3.3.1	Leikkausmatriisien rakenne	16
3.3.2	Predikaatit joukossa {T, F, *}	18
3.3.3	Käänteispredikaatit	19
3.3.4	Predikaatit joukossa {0, 1, T, F, *}	20
4	PostGIS	21
4.1	Spatiaaliset tietokannat ja PostGIS	21
4.2	PostGIS kehittyä	22
4.3	Spatiaalinen indeksi ja GiST	23
4.4	Kyselyiden rakenne	25
4.5	PostGIS-funktiot ja -operaattorit	27
5	Laadunarvioinnin menetelmät ja työkalut projektissa	31
5.1	Tavoitteet ja lähtötilanne	31
5.2	Laadunarviointi ELF-projektissa	31
5.2.1	Aineistotuotteet ja -teemat	33
5.2.2	Laatusäännöt	34
5.3	Kehitysympäristö ja työkalut	36
5.3.1	Aineiston lataaminen tiedostoksi WFS-palvelusta	36
5.3.2	QGIS aineiston tarkastelussa ja prosessoinnissa	37
6	Laatutestit PostGIS-kyselyinä	39
6.1	Täydellisyyden testaus	39
6.1.1	Kohteiden minimikoko	39
6.1.2	Pakollisten kohteiden läsnäolo	40
6.2	Käsitteellisen eheyden testaus	41
6.3	Arvojoukkoeheyden testaus	42
6.4	Formaattieheyden testaus	42
6.5	Topologisen eheyden testaus	42
6.5.1	Viivojen yhdistyvyys	43
6.5.2	Polygonien väliset aukot ja säleet	43
7	Johtopäätökset	50
	Lähdeluettelo	53
	Liite 1 – ELF-laatusäännöt ja laatutesteissä käytetyt PostGIS-funktiot	

Merkinnät

∂	topologinen reuna
\cap	leikkaus
\cup	unioni
\emptyset	tyhjä joukko
\wedge	AND, looginen operaattori
\vee	OR, looginen operaattori
\neg	NOT, looginen operaattori
\Leftrightarrow	yhtä kuin
\Rightarrow	looginen johtopäätös, seuraava on seurausta edellisestä
\in	kuuluu joukkoon

Termit ja lyhenteet

Aineisto	Yksilöitävissä oleva kokoelma tietoja, synonyyminä TSK 45:n (2014) mukaiselle tietoaineisto-termille.
Avoin lähdekoodi	Lähdekoodi, jonka lisenssi sallii vapaan käytön, kopioinnin, levityksen ja muokkauksen (engl. open source) (JHS 169 2012).
ELF	European Location Framework.
ESDIN	European Spatial Data Infrastructure with a Best Practice Network.
INSPIRE	Infrastructure of Spatial Information in Europe.
JHS	Julkisen Hallinnon Suositus.
KMTK	Kansallinen maastotietokanta.
Kohde	Objekti, joka vastaa yksilöitävissä olevaa reaalimaailman abstraktia tai konkreettista asiaa tai ilmiötä (TSK 45 2014).
Käytettävyys	Asian tai esineen helppokäyttöisyys tietyn tavoitteen saavuttamisessa.
Leikkaus	Joukko, johon kuuluvat kaikki joukon A elementit, jotka esiintyvät myös joukossa B.
Objektiivinen laatu	Virheiden määrään perustuva laadun käsite.
Paikkatieto	Tieto kohteista, joiden paikka Maan suhteen tunnetaan (TSK 45 2014).
Tyhjä joukko	Joukko, joka ei sisällä yhtään alkia.
SRID	Spatial Reference System Identifier, eli koordinaattijärjestelmän tunnus.
Subjektiiivinen laatu	Mielikuvaan perustuva laadun käsite.
Sulkeuma	Topologisesti rajattu joukko pisteitä, jotka muodostavat yhtenäisen alueen.
Unioni	Joukkojen A ja B yhdiste, johon kuuluvat kaikki niissä esiintyvät, erilaiset alkio.
WFS	Web Feature Service (tietokohdepalvelu; ISO 19142). Latauspalvelu vektoriaineistolle.

1 Johdanto

1.1 Työn tausta

Paikkatieto on tietoa, johon liittyy sijainti. Paikkatiedoilla kuvataan luonnollisia tai rakennettuja kohteita Maan pinnalla, mutta paikkatietoa on käytännössä kaikki asiat tai ilmiöt, joiden sijainti tunnetaan. Ihminen on kautta aikojen pyrkinyt jäljentämään todellisuutta eri muodoissa pystyäkseen hahmottamaan ympäristöään paremmin ja laajemmin sekä kommunikoimaan siitä. Paikkatiedot ovat muodostuneet elintärkeäksi resurssiksi ihmiskunnan kehityksen ja toiminnan kannalta, ja ongelmat niiden laadussa ikään kuin vääristävät kuvausta todellisuudesta. Laatu muodostuu useista tekijöistä monimutkaisten tuotantoprosessien eri vaiheissa, ja parhaiten siihen voidaan vaikuttaa tuotantoprosesseja kehittämällä.

Paikkatietojen automaattinen laadunarviointi on menetelmä, jonka avulla paikkatietojen eheys voidaan varmistaa nopeasti ja koko tietokannan laajuisesti. Menetelmää voidaan hyödyntää kaikissa tietoaineiston elinkaaren vaiheissa, tuotannosta ylläpitoon. Ohjelmallisesti tehtävät tarkastukset perustuvat tietoaineistojen vaatimusmäärittelyihin. Vuosien 2008–2011 toiminnassa olleessa ESDIN-projektissa (European Spatial Data Infrastructure Network) esitettiin periaate automaattiselle laadunarvioinnille, joka tutkii aineiston sisäistä laatua ISO 19100 -sarjan standardeihin perustuvan laatumallin pohjalta. (Beare et al. 2010.) Projektissa todettiin, että automaattinen laadunarviointi on tunnetuista laadunarviointimenetelmistä kattavin ja kustannustehokkain (Overton 2011).

Automaattisessa laadunarvioinnissa on kyse koko tietokannan laajuisesti suoritettavasta, paikkatietoanalyysistä, jossa testataan, noudattaako aineisto sille määriteltäjä sääntöjä. Automaattisuus poistaa arviointiprosessista ihmisen tekemät virheet ja samalla siirtää valtaosan työmäärän tietokoneen prosessoitavaksi. Silmämääräisesti useimpia geometrisiä virheitä on käytännössä mahdotonta löytää jo suurten aineistomäärien takia. Menetelmän avulla voidaan tutkia sellaisia aineistoissa esiintyviä virheitä, joiden löytäminen ei vaadi aineiston ulkopuolista referenssitietoa. (Beare et al. 2010.)

Eurooppalaiset maanmittauslaitokset ovat luomassa European Location Framework (ELF) -projektissa osallisten maiden kansallisista aineistoista rajojen yli harmonisoitua viitekehystä. Projektissa on tarkoitus tuoda INSPIRE-yhteensopivia aineistoja paremmin saataville. Laatu on eräs avainelementeistä lukuisten tuottajaorganisaatioiden aineistoja yhteensovittaessa, sillä eri organisaatioiden tuotantoprosessit ja toteutukset saattavat poiketa toisistaan ratkaisevasti. Poikkeavan huonolaatuinen datasetti heikentää koko aineiston käytettävyyttä vaikuttamalla negatiivisesti paikkatietoanalyysiin. ELF-projektissa automaattisen laadunarvioinnin avulla varmistetaan, että yhteiseen tietoaineistoon liitettävät aineistot ovat määrittelyiden mukaisia. Monien tuottajaorganisaatioiden tapauksessa aineistojen tuotantoprosesseja ei varsinaisesti voida hallita suoraan, joten yhteiset säännöt ovat hyvin tärkeitä. Tuottajaorganisaatioiden pitää olla kykeneviä toimittamaan riittävän laadukkaita aineistoja.

Paikkatiedon laatua on tutkittu melko paljon niin Suomessa kuin kansainvälisesti. Paikkatietoaineistojen laadunhallintaa varten on saatavilla kansainvälisiä ISO 19100 -sarjan standardeja, ja Suomessa niiden pohjalta laadittu JHS 160 -suositus. Paikkatietojen laatua on testattu paikkatieto-ohjelmistojen avulla, mutta automaattiseen laadunarviointiin keskittyneitä ohjelmistoja on vielä tarjolla vain joitain.

Paikkatiedon laatu ei ole asia, jonka pitäisi estää sen saatavuus, sillä mitään paikkatietoa ei voida pitää täydellisenä kuvauksena todellisuudesta. Vaikka paikkatiedot olisivat runsaita ja yksityiskohtaisia, ne eivät voi olla täydellisiä. Laadunarvioinnin tarkoituksena on tarjota tietoa näistä epätäydellisyyksistä, jotta paikkatiedon käyttäjä voi päätellä omalla kohdallaan, voidaanko aineistoa käyttää aiottuun tarkoitukseen. (Beare et al. 2010.)

1.2 Tutkimuksen tavoitteet ja rajaukset

Työn ensisijaisena tavoitteena on selvittää, voidaanko ELF-projektin laatuvaatimusmäärittelyiden mukainen paikkatietojen automaattinen laadunarviointi toteuttaa avoimen lähdekoodin ohjelmistoilla. Työn käytännön osuus toteutettiin pääasiassa PostGIS-tietokantaohjelmiston SQL-kyselyinä, joiden oli tarkoitus kattaa kaikki projektin laatuvaatimusmäärittelyt.

Vaatimusmäärittelyt ovat hyvin tärkeä osa laadunarviointia, koska laatua testataan juuri näitä määrittelyitä vastaan. Laatuvaatimusten muodostamiseen ei työssä puututa, mutta työn teoriaosuudessa pyritään tarjoamaan riittävä ymmärrys vaatimusmäärittelyiden sisällöstä, jotta käytännön osuudessa valmiiden määrittelyiden pohjalta tehtyjä testejä olisi mahdollista ymmärtää.

Työssä käsitellään ainoastaan vektoriaineiston laatua. Rasterimuotoisen aineiston laadunarviointimenettely poikkeaa tästä tapauksesta merkittävästi, joten tässä työssä esitetyt menetelmiä ei voida siihen soveltaa.

Paikkatiedon laatu heikkenee tuotantoprosessin aikana muodostuneiden virheiden takia. Työssä keskitytään automaattisessa laadunarvioinnissa käytettyihin menetelmiin ja tietoihin. Muita laadunhallinnan aspektejä ei käsitellä tarkemmin, koska ne eivät liity suoraan tutkimuskysymyksen ratkaisuun.

Käytännön osuudessa yksittäisen PostGIS-kyselyn tavoitteena on palauttaa datasetti, joka sisältää kaikki laatuvaatimusta rikkovat kohteet. Jos mahdollista, pyritään laatuvaatimusta vastaavaa ongelmaa käsittelevän lähteen käyttöön. Suoritusaikoihin tai koodin tyyliin ei ole tässä vaiheessa tarkoitus kiinnittää erityistä huomiota, vaan olennaista on PostGIS-funktioiden hyödyntäminen ongelman ratkaisemiseksi.

1.3 Työn rakenne

Luku 2 esittelee paikkatiedon laadun perusteita. Aluksi kerrotaan taustatietoja laadusta ja paikkatiedoista sekä paikkatietojen laadunhallintaa koskevista standardeista. Luvun loppu on omistettu laadunarvioinnille. Laatutekijät ovat ensisijaisen tärkeitä työn tulosten ymmärtämiseksi. Laatutekijöiden lisäksi kerrotaan yleisesti laatumittareista sekä laadunarviointimenetelmistä pääasiassa ISO 19157 -standardiin sekä JHS 160 suosituksen viitaten. Lopuksi kerrotaan saatavilla olevista laadunarviointiohjelmistoista.

Luvussa 3 käydään läpi spatiaalisten relaatioiden teoria. Spatiaaliset relaatiot ovat työn käytännön osuuden kannalta hyvin olennaista teoriaa. PostGIS-ohjelmiston paikkatietofunktiot perustuvat DE-9IM-relaatioihin, joita luku käsittelee melko syvällisesti.

Luvussa 4 kerrotaan käytännön osuudessa hyödynnetystä PostGIS-tietokantaohjelmistosta. Aluksi käydään läpi spatiaalisten tietokantojen perusteita ja PostGIS:n historiaa. Seuraavaksi kerrotaan, miten paikkatietokannoissa käytetään koordinaattijärjestelmiä,

jonka jälkeen keskitytään kyselyiden toimintaa merkittävästi nopeuttavan spatiaalisen indeksin perusteisiin. Kahdessa viimeisessä aliluvussa esitetään aluksi PostgreSQL-kyselyiden perusrakenne, jonka jälkeen käydään läpi työssä käytettyjä PostGIS-laajennuksen funktioita ja operaattoreita.

Luku 5 esittelee tarkemmin työssä käytetyt aineistot sekä laadunarvioinnin menetelmät ja työkalut. Aluksi kerrotaan ELF-projektista, testattavista aineistoista sekä laatuvaatimusmäärittelyistä, joiden perusteella testit laaditaan. Lopuksi kerrotaan testien laatimisessa käytetyistä työkaluista.

Luvussa 6 esitellään työn varsinainen käytännön osuus, eli laatutestit PostGIS-kyselyinä. Kyselyistä annetaan koodiesimerkkejä, joiden toiminta käydään perusteellisesti läpi. Kyselyt on jaettu laatulementeittäin. Työn tulokset esitellään liitteenä, jossa on osa alkupe-
räisestä määrittelydokumentista ja testeissä käytetyt PostGIS-funktiot.

Luvussa 7 tehdään johtopäätöksiä työn tulosten perusteella ja arvioidaan jatkokehitysmahdollisuuksista.

2 Paikkatiedot ja niiden laatu

Tässä luvussa perehdytään työn taustatietoihin ja keskeiseen teoriaan alan kirjallisuuden sekä joidenkin havainnollisten esimerkkien johdattamana. Luku tarjoaa tietoja laadusta, paikkatiedoista itsestään, sekä erityisesti automaattisessa laadunarvioinnissa tarpeellisesta paikkatietojen laadun teoriasta, jossa keskitytään erityisesti laatutekijöihin.

2.1 Laatu käsitteenä

Laatu on käsitteenä jokaiselle tuttu, mutta sen merkitystä voi olla vaikeaa todeta yksiselitteisesti. Määritelmiä sille on useita, mutta ISO 9000 (2005) -standardissa laatu määritellään asteena, jolla jonkin asian luontaiset ominaisuudet täyttävät sille asetetut vaatimukset. Laatua ei siis voida arvioida, ennen kuin tarkasteltavalle asialle on tunnistettu sen hyödyntämisen kannalta tärkeimmät ominaisuudet ja määritelty niille saavutettavissa olevat tavoitteet.

Laatu voidaan tulkita myös objektiivisena tai subjektiivisena laatuna. Objektiivinen laatu käsittelee jonkin asian mitattavia ominaisuuksia, eli objektiivisesti laadukkaassa asiassa virheen määrä on vähäinen. Subjektiivisesti, eli kokemuseräisesti laadukas asia vastaa käyttäjän tarpeita hyvin. Paikkatiedoissa aineiston laatu on mitattavissa virheellisten kohteiden lukumäärän perusteella, mistä syystä se voidaan tulkita objektiiviseksi laaduksi. Lopputuotteiden, kuten karttojen, mallien tai palveluiden käytettävyys voitaisiin puolestaan käsittää subjektiivisena laatuna. (Devillers & Jeansoulin 2006.)

Paikkatiedoissa laatu voi ilmetä puuttuvina tai väärin sijoitettuina kohteina, ja niiden ominaisuustiedot voivat olla vääriä tai niitä ei ole lainkaan. Käyttäjän kannalta tällaisilla ongelmilla voi olla suurikin merkitys. Käyttäjälle tulisikin tarjota kuvaus laadutiedoista, jotta aineiston ajantasaisuus sekä virheiden mahdollisuus tulisivat mahdollisimman hyvin selville. (Devillers & Jeansoulin 2006.) Laatua voidaan lähestyä tietojen näkökulmasta, jolloin sitä mitataan aineistoissa esiintyvänä virheinä (ISO 19157 2013). Organisaatioiden näkökulmasta laatu voidaan mitata organisaation kykynä tuottaa haluttua laatua (ISO 19158 2012). Kolmantena näkökulmana voidaan pitää jo aiemmin mainittua käyttäjälähtöistä, subjektiivista näkökulmaa.

2.2 Paikkatietojen olemus ja virheet

Paikkatieto on tietoa kohteista, joiden sijainti Maan suhteen tunnetaan. Tyypillisesti paikkatieto kuvaa luonnollisen tai rakennetun ympäristön kohteita, mutta mikä tahansa asia, jonka sijainti voidaan muodollisesti määrittää, on paikkatietoa. Paikkatieto voidaan jakaa kahteen päätyyppiin, sijaintitietoon sekä ominaisuustietoon. Sijaintitiedolla tarkoitetaan kohteen fyysisestä sijainnista kertovaa tietoa, esimerkiksi koordinaatteja. Ominaisuustiedot ovat monenlaisia, kohteen luonteesta kertovia tietoja. Näitä voivat olla esimerkiksi paikannimet, osoitteet, maankäytön tyyppi, asukasmäärät tai mikä tahansa muu tieto, jolla kohdetta voidaan kuvailla. (TSK 45 2014.)

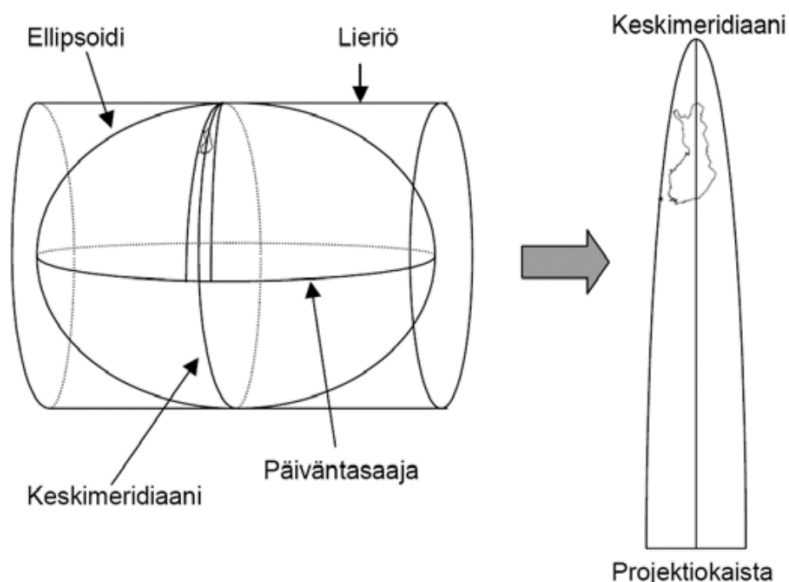
Paikkatietoja tuottavat ja hallitsevat lukuisat organisaatiot ympäri maailman. Suomessa julkisia paikkatietoja hallinnoivia viranomaisia ovat muun muassa Maanmittauslaitos, Liikennevirasto, kunnat, Ilmatieteen laitos, Suomen ympäristökeskus, ELY-keskukset, Geologian tutkimuskeskus, Tilastokeskus ja puolustusvoimat. (922/2014 2014). Euroopassa paikkatietojen yhteiskäytöstä on laadittu EU:n laajuinen INSPIRE-direktiivi, jonka

avulla eri maiden kansalliset aineistot on tarkoitus saada helpommin käyttöön. Saatavuuden helpottamisen lisäksi yhteiskäytön edistäminen vähentää merkittävästi päällekkäistä työtä tiedonkeruussa. (2007/2/EY 2007.)

Paikkatieto jaetaan karkeasti vektori- ja rasteriaineistoihin. Rasteriaineistolla tarkoitetaan tasakokoisista yksiköistä, eli pikseleistä koostuvaa aineistoa. Digitaaliset kuvat muodostuvat pikseleistä, ja rasteriaineistolla tarkoitetaan paikkatietojen yhteydessä useimmiten erilaisia ilma- tai satelliittikuvia. Vektoriaineistossa paikkatietokohteita kuvataan pisteiden, viivojen tai polygonien muodossa tai näiden yhdistelminä. (TSK 45 2014.) Vektoriaineistoa tuotetaan yleisemmin digitoimalla, eli jäljentämällä todellisen maailman kohteita ilmakuvista vektorimuotoon. Ilmakuvien digitointiin käytetään yleisemmin kuvaparien kolmiulotteisen tarkastelun mahdollistavaa stereotyöasemaa. (Devillers & Jeansoulin 2006.)

2.3 Koordinaatit paikkatietojärjestelmissä

Koordinaattijärjestelmiä (engl. coordinate reference system, CRS) tarvitaan, jotta tiedetään, miten paikkatieto ilmenee suhteessa Maahan. Paikkatietojen koordinaatit pitää aina olla määritelty jossain koordinaattijärjestelmässä. Koordinaattijärjestelmä koostuu kahdesta komponentista, datumista ja koordinaatistosta. Datumi kiinnittää koordinaatiston suhteessa maahan, määrittäen origon paikan sekä koordinaatiston asennon. Koordinaatisto määrittää koordinaattijärjestelmän mitta-asteikon. Maapallo ei ole litteä, joten kaksikulotteiset koordinaatit projisoidaan yleensä lieriölle tai tasolle. Esimerkiksi kuvan 5 Transverse Mercator (TM) -projektiossa Maan pinta kuvataan kaistoina vaakasuuntaisen sylinterin pinnalle, jonka keskikohta leikkaa Maan pintaa. Kaistat ovat litteitä, joten mittakaavavirhe joudutaan ottamaan havainnoissa huomioon. Jos kaistan keskikohta on maan alla, on mittakaavavirhe suurin keskikohdassa. Jos keskikohta on puolestaan tangentiaalisesti maan päällä, on mittakaavavirhe suurin reunoissa. (Vermeer 2015.)



Kuva 5. Transverse Mercator -projektio (Häkli et. al 2009).

Eri koordinaattijärjestelmissä olevia aineistoja voidaan verrata ja käyttää yhdessä koordinaattimuunnosten avulla. Koordinaattijärjestelmien väliset muunnokset ovat yksinkertaisia, sillä niihin tarvitaan vain koordinaatiston origon siirto ja kierto. Koordinaatistoa tai karttaprojektiota muunnettaessa joudutaan huomioimaan useampia parametreja. Helmert-

muunnos on yleisin tasomuunnos, ja se voidaan suorittaa kaavan 15 mukaisesti. Kaavassa $1+m$ on mittakaavakerroin, R on rotaatiomatriisi, ja t siirtovektori. Rotaatiomatriisissa e_x , e_y ja e_z ovat pieniä kiertokulmia akseleidensa ympäri. (Vermeer 2015.)

$$\begin{bmatrix} x_2 \\ y_2 \\ z_2 \end{bmatrix} = (1 + m)R \begin{bmatrix} x_1 \\ y_1 \\ z_1 \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \\ t_z \end{bmatrix} \quad (15)$$

, jossa

$$R = \begin{bmatrix} 1 & e_z & -e_y \\ -e_z & 1 & e_x \\ e_y & -e_x & 1 \end{bmatrix}$$

Paikkatietojärjestelmissä ja spatiaalisissa tietokannoissa tietoaineistoille määritellyt koordinaattijärjestelmät merkitään SRID-tunnuksin (engl. Spatial Referencing System Identifier). PostGIS:ssä SRID-tunnuksena käytetään nelinumeroisia European Petrol Survey Groupin määrittelemiä EPSG-koodeja. (PostGIS 2015b.) Aineistoissa voidaan käyttää mitä tahansa alueelle soveltuvaa koordinaattijärjestelmää. Suomalaisissa aineistoissa käytetään useimmin ETRS-TM35FIN-koordinaattijärjestelmää, jonka EPSG-tunnus on 3067 (JHS 154 2012). Euroopan alueella on käytössä ETRS89-GRS80-koordinaattijärjestelmä, jonka EPSG on 4258. ITRS, jonka vuoden 2008 realisaation ITRF2008:n EPSG on 5332, on käytössä kansainvälisesti. Myös korkeuskoordinaateille on saatavilla omat koordinaattijärjestelmänsä, kuten myös yhdistelmäkoordinaateille, joissa ilmoitetaan samalla EPSG-tunnuksella 2D:n sekä korkeuden koordinaattijärjestelmä. (INSPIRE 2009.)

2.4 Standardit ja suositukset

Suomessa paikkatiedon laadunhallintaa varten on saatavilla julkisen hallinnon JHS-suositus 160, joka perustuu pääosin ISO 19113 Geographic information – Quality principles, ISO 19114 Geographic information – Quality evaluation procedures ja ISO/TS 19138 Geographic information – Data quality measures -standardeihin sekä yleisesti laadunhallintaa käsittelevään ISO 9000 -standardisarjaan (JHS 160 2006, s. 2). JHS 160 tiivistää paikkatiedon laadunhallinnan käsitteet selkeäksi suomenkieliseksi kokonaisuudeksi. (JHS 160 2006.)

Kansainvälisistä standardeista paikkatiedon laadun kannalta olennaisimmat ja viimeisimmät standardit ovat ISO 19157 ja 19158. ISO 19157 Geographic information – Data quality yhdistää aiempia ISO-laatuststandardeja laajemmaksi ja uudistetuksi kokonaisuudeksi. ISO/TS 19158 Geographic information – Quality assurance of data supply käsittelee puolestaan paikkatiedon tuottajan valmiutta tuottaa tarpeeksi laadukasta aineistoa. (ISO 19157 2013, ISO/TS 19158 2012.)

2.5 Laatutekijät

Paikkatiedon laatua arvioidaan laatutekijöiden perusteella. Laatutekijät on ryhmitelty paikkatietojen laadun eri näkökulmia edustaviin kategorioihin. Laatutekijöiden avulla voidaan esittää, miten hyvin paikkatietoaineistot vastaavat aineistomäärittelyitään. Laadunarvioinnin tuloksena saadaan tietoa virheiden määristä. Laatutekijöiden laatua voidaan kuvata metalaadun avulla. (ISO 19157 2013.) Automaattisen laadunarvioinnin avulla voidaan arvioida niistä loogista eheyttä ja täydellisyyttä.

Laatutekijät muodostetaan aineiston laatuvaatimusten pohjalta kuvaamaan yksittäisiä laadunarvioinnin kohteita. Laatutekijät jaetaan ISO 19113 -standardissa sekä JHS 160 -suosituksessa mitattaviin ja kuvaileviin laatutekijöihin, joista mitattavat laatutekijät jakautuvat edelleen laatutekijöiden osatekijöihin. Uudistetussa ISO 19157 -standardissa ei käytetä tällaista jaottelua. Laatutekijöihin ISO 19157:ssa on sen sijaan lisätty käytettävyyselementti. JHS 160:n ja ISO 19157:n yhtenevät laatutekijät esitetään taulukossa 1. (JHS 160 2006, ISO 19157 2013.)

Täydellisyys (engl. completeness) kuvaa aineiston eri elementtien olemassaoloa tai puuttumista. Tämän kategorian laatutekijöitä ovat ylimääräinen (engl. commission) ja puuttuva tieto (engl. omission). (JHS 160 2006 s. 14, ISO 19157 2013.)

Looginen eheys (engl. logical consistency) kuvaa elementtien sääntöjenmukaisuuden asetta aineiston rakenteeseen, ominaisuuksiin ja keskinäisiin relaatioihin nähden. Laatutekijöistä käsitteellinen eheys (engl. conceptual consistency) liittyy käsiteskeeman sääntöjenmukaisuuteen. Arvojoukkoeheys (engl. domain consistency) kertoo, kuuluvatko aineiston eri elementtien arvot hyväksyttävyn arvojoukkoon. Formaattieheys (engl. format consistency) ilmaisee suhteen, jolla aineiston rakenne on yhtenevä fyysisen tietomallin kanssa. Topologinen eheys (engl. topological consistency) kertoo, onko aineiston geometrian suhteet topologisesti sääntöjenmukaisia. (JHS 160 2006, ISO 19157 2013.)

Sijaintitarkkuus (engl. positional accuracy) kuvaa, miten tarkasti sijainti on määritetty. Laatutekijöistä absoluuttinen tai ulkoinen sijaintitarkkuus (engl. absolute or external positional accuracy) kertoo, ovatko havaitut ja todelliset koordinaatit yhteneviä. Rasteritiedon sijaintitarkkuus (engl. gridded data positional accuracy) kertoo pikseleiden läheisyydestä niiden esittämien todellisten kohteiden sijaintiin. Suhteellinen sijaintitarkkuus (engl. relative internal positional accuracy) kertoo kohteiden suhteellisen sijainnin tarkkuudesta todellisten kohteiden suhteellisiin sijainteihin nähden. (JHS 160 2006, ISO 19157 2013.)

Ajallinen tarkkuus (engl. temporal accuracy) on tarkkuutta kohteiden aikaan liittyvissä ominaisuuksissa. Laatutekijöistä ajan mittauksen tarkkuus (accuracy of a time measurement) tarkoittaa nimensä mukaisesti ajan mittaamisen tarkkuutta. Ajallinen eheys (temporal consistency) kertoo ajallisten järjestyksen todenmukaisuudesta. Ajanmukaisuus (engl. temporal validity) puolestaan aikahavaintojen todenmukaisuudesta. (JHS 160 2006, ISO 19157 2013.)

Temaattinen tarkkuus (engl. thematic accuracy) kertoo määrällisten attribuuttien tarkkuudesta sekä laadullisten attribuuttien oikeellisuudesta. Laatutekijöistä luokittelun oikeellisuus (engl. thematic classification correctness) kertoo kohteiden luokittelun tarkkuudesta. Laadullisen ominaisuustiedon oikeellisuus (engl. non-quantitative attribute accuracy) kertoo sellaisen ominaisuustiedon oikeellisuudesta, mitä ei voida mitata. Määrällisen ominaisuustiedon tarkkuus (engl. quantitative attribute accuracy) kertoo mitattavien ominaisuustietojen tarkkuudesta. (JHS 160 2006, ISO 19157 2013.)

Käytettävyystekijä ISO 19157:ssa perustuu käyttäjien asettamille vaatimuksille aineiston laadusta. Mikä tahansa laatutekijä sopii käytettävyyden arviointiin, mutta kuten kuvailevien laatutekijöiden yhteydessä todettiin, joitain vaatimuksia ei voida näiden tekijöiden perusteella arvioimaan. Käytettävyystekijän avulla voidaan esimerkiksi kertoa, miten hyvin aineisto sopii johonkin tiettyyn käyttötarkoitukseen. (ISO 19157 2013.)

Taulukko 1. Laatutekijät JHS 160 -suosituksessa sekä ISO 19157 -standardissa.

JHS 160		ISO 19137	
Tekijä	Osatekijät	Category	Data quality element
Täydellisyys	Ylimääräinen tieto Puuttuva tieto	Completeness	Completeness commission Completeness ommission
Looginen eheys	Käsitteellinen eheys Arvojoukkoeheys Formaattieheys Topologinen eheys	Logical consistency	Conceptual consistency Domain consistency Format consistency Topological consistency
Sijaintitarkkuus	Absoluuttinen sijaintitarkkuus Rasteritiedon sijaintitarkkuus Suhteellinen sijaintitarkkuus	Positional accuracy	Absolute external accuracy Gridded data accuracy Relative internal accuracy
Ajallinen tarkkuus	Ajan mittauksen tarkkuus Ajallinen eheys Ajanmukaisuus	Temporal accuracy	Accuracy of a time measurement Temporal consistency Temporal validity
Temaattinen tarkkuus	Luokittelun oikeelli- suus Ei-kvantitatiivisen ominaisuustiedon oikeellisuus Kvantitatiivisen ominaisuustiedon tarkkuus	Thematic accuracy	Thematic classification correctness Non-quantitative at- tribute accuracy Quantitative attribute accuracy

Kuvailevat laatutekijät eivät kuulu enää ISO 19157 -standardissa laatutekijöihin. Ne edustavat JHS 160:ssa sellaisia laatutietoja, joita ei pystytä mittaamalla arvioimaan. Niillä ei ole varsinaisia osatekijöitä, vaikka historiatiedot jaetaan kahteen osaan. Historiatiedot kertovat aineiston tuotannon prosessien taustoista. Prosessointihistoria on tietoa itse aineistolle tehdyistä toimenpiteistä. Alkuperätiedot puolestaan kertovat, mistä aineisto on peräisin. Käyttötarkoitus kertoo, mihin käyttöön aineisto on alkujaan tarkoitettu. Käyttökokemukset ovat tietoa aineiston käyttöhistoriasta sekä tuottajan että loppukäyttäjän toimesta. (JHS 160 2006.) ISO 19157 -standardissa vastaavia laatuun vaikuttavia, ulkoisia asioita voidaan raportoida osana metadataa ISO 19115 -standardin mukaisesti epäsuoran laadunarvioinnin yhteydessä (ISO 19157 2013, s. 23).

2.6 Laatumittarit ja laadunarviointimenetelmät

Laatumittareita tarvitaan eri aineistojen keskinäisen laadun vertailuun. Laatumittari voi olla esimerkiksi virheellisten kohteiden osuus aineistossa prosentteina. Mittareiden tulisi olla riittävän selkeitä ja muodoltaan yhtenäisiä. Toisinaan riittävän hyvää kuvaa laadusta ei pystytä tarjoamaan yhdellä mittarilla, jolloin samalle perusjoukkoille joudutaan määrittelemään useampia mittareita. Laatumittarilla saadaan tuotettua laatutulos, joita voidaan antaa jokaiselle laatulementille useita. Laatutulokselle määritellään tietotyyppi, kuten

Kaikille laatumittareille pitää erikseen määritellä omat arviointimenetelmänsä. Laadunarviointimenetelmät jaetaan epäsuoriin ja suoriin laadunarviointimenetelmiin. Laadunarvioinnissa verrataan aineiston todellista laatua laatuvaioitteisiin. Laadunarviointi tehdään joko näyte- tai kokonaistutkimuksena. Kokonaistutkimuksessa tarkasteltavaan perusjoukkoon kuuluu jokainen aineiston kohde. Näytetarkastuksessa perusjoukko muodostetaan poimimalla aineistosta otanta todennäköisyysteorian mukaisesti. Otannasta lasketun laatu tuloksen perusteella perusjoukko voidaan joko hyväksyä tai hylätä. (JHS 160 2006, ISO 19157 2013.)

```
graph TD; A[Paikkatiedon laatu] -.->|Laatukuvauksen laajuus| B[Laatukuvauksen laajuus]; B -->|jakautuu| C[Laatutuloksen laajuus]; A -->|ilmaistaan| D[Laatutekijä]; A -.->|raportoidaan| E[Laaturaportti]; A -.->|raportoidaan| F[Metadata ISO 19115]; D -->|kuvaillaan| G[Laatumittari]; D -->|kuvaillaan| H[Laadunarviointi]; D -->|kuvaillaan| I[Laatutulos]; D -->|kuvaillaan| J[Metalaatu];
```

Kuva 1. Paikkatiedon laadun käsitteet (mukaillen ISO 19157).

9

Ominaisuustarkastuksessa perusjoukon kohteet luokitellaan virheellisiksi tai virheettömiksi vertaamalla niitä vaatimusmäärittelyihin. Perusjoukko voidaan hylätä tai hyväksyä virheellisten kohteiden lukumäärän perusteella. (JHS 160 2006, liite 4.)

Muuttujatarkastuksessa kohteista mitataan jokin suure laatuominaisuudeksi. Perusjoukon tai otannan hyväksymiskriteeri muodostetaan perustuen normaalisti jakautuneiksi oletettujen mittaustulosten jakauman keskiarvon tai -hajonnan estimaatteihin. Perusjoukko tai otos voidaan hyväksyä tai hylätä vertaamalla hyväksymiskriteeriä tarkastusrajoihin. Tarkastusrajalle määritellään yläraja U sekä mahdollisesti myös alaraja L. Kohde on virheellinen, jos sen mitattu laatuominaisuus on suurempi kuin U tai pienempi kuin L. (JHS 160 2006, liite 4.)

Epäsuorat laadunarviointimenetelmät perustuvat aineiston ulkopuolisten tietojen hyödyntämiseen. Ulkopuolisten tietojen käyttö tässä tapauksessa voi tarkoittaa muun muassa asiantuntija-arvioiden tai muiden kuvailevien laatatietojen käyttöä. Epäsuoria arviointimenetelmiä tulisi käyttää vain sellaisissa, harvoissa tapauksissa, joissa suoraa arviointimenetelmää ei voida lainkaan hyödyntää. (JHS 160 2006, ISO 19157 2013.)

Laatu raportoidaan osana metadataa ISO 19115 -standardin mukaisesti. Lisätietoja laadusta voidaan tarjota vapaamuotoisena laaturaporttina, mutta sen ei ole tarkoitus korvata metatietoja laadusta. (ISO 19157 2013.)

2.7 Metalaatu

Metalaadun avulla voidaan kuvata laadunarvioinnin tai laatur tuloksen laatua. Metalaatutekijät ovat määrällisiä tai laadullisia lausuntoja, joissa arvioidaan laatumittarin ja laatur tuloksen sekä arviointimenetelmien soveltuvuutta käyttötarkoituksiinsa. (ISO 19157 2013.)

Luotettavuus kertoo laatur tuloksen luotettavuudesta. Luotettavuuden avulla arvioidaan sovellettavien arviointimenetelmien luotettavuutta aineiston laadun arvioinnissa. (ISO 19157 2013.)

Edustavuus kertoo, missä laajuudessa aineiston laatua on arvioitu. Aineistosta voidaan arvioida erilaisia osa-alueita tai koko aineistoa. Tärkeintä on tuoda esille, missä määrin arvioinnin kohde edustaa aineiston nykytilaa. (ISO 19157 2013.)

Homogeenisuus kertoo laatur tulosten yhdenmukaisuudesta. Yhdenmukaisuutta arvioidaan vertaamalla otantojen laatur tuloksia toisiinsa. Homogeenisuus voidaan määrittää siis vain, jos aineiston laatua on arvioitu otoksina. Homogeenisuutta arvioidaan usein, kun kyseessä on eri toimijoiden tekemät tai eri ajankohdilta olevat aineistot. (ISO 19157 2013.)

2.8 Laatusäännöt

ELF-projektissa laatusäännöt ovat aineistojen käyttötapauksiin, joihinkin INSPIREN eheyssääntöihin sekä maanmittauslaitosten mahdollisesti tarjoamiin aineistoihin perustuvia laatuvaatimusmäärittelyitä. Sääntöihin kuuluu ainoastaan automaattisesti testattavissa olevia kriteereitä, ja ne keskittyvät edistämään aineistojen harmonisointia tunnistamalla virheitä sen topologisessa rakenteessa sekä varmistamalla kohteiden ja niiden ominaisuuksien säännönmukaisuus aineistomäärittelyihin nähden. (Werhahn et al. 2015, Annex A.)

Harmonisointia varten laadittujen sääntöjen tulee olla yksiselitteisiä ja kaikkien ymmärrettävissä etenkin kansainvälisessä, useita toimijoita kattavassa organisaatiossa. Yksiselitteisyyttä on pyritty edistämään käyttämällä sääntöjen määrittelyissä yhtenäistä rakennetta, RuleSpeakia. (Werhahn et al. 2015, Annex A.)

RuleSpeakin tarkoitus on, että eri toimijoiden olisi helpompi ilmaista samoja asioita samoilla tavoin. RuleSpeakin lauserakenteet ovat kaavoja tai malleja, joilla voidaan luonnollisella kielellä ilmaista asioita johdonmukaisesti ja hyvässä järjestyksessä. Lauserakenteet on tarkoitettu ohjeistusten ja kommunikaation selkeyttämiseen, eikä niinkään teknisten implementaatioiden määrittelyyn. (Ross 2009.)

RuleSpeak-säännöt ovat määrääviä tai ohjeellisia. Määrääviin sääntöihin pitää kuulua ”must”- tai ”only”-avainsana. Vastaavasti ohjeellisiin sääntöihin pitää kuulua ”may”- tai ”need not”-avainsana. Ohjeellisissa säännöissä ei voida käyttää määrääviä avainsanoja. Jos määräävää avainsanaa käytetään ohjeellisessa säännössä, säännöstä tulee määräävä. (Ross 2009.) Avainsanaa ”the following” käytetään aloittamaan listaus. Avainsanan jälkeen ennen varsinaista listaa määritellään, kuinka monen listan kohdan pitää päteä. (Ross 2013.)

Taulukko 2. Avainsanoja RuleSpeak-säännöissä.

Avainsanat	Kuvaus
”must”	Pakollisuus
”must not”	Kielto
”must be computed as”	Laskentamenetelmän määrittely
”must be considered ... if”	Jotain asiaa pitää harkita, kun ehto täyttyy
”must be performed when”	Jokin asia pitää suorittaa, kun ehto täyttyy
”may”	Jokin asia on sallittu, mutta ei pakollinen
”may ... only”	Jokin asia on sallittu vain tietyin ehdoin
”need not”	Jokin asia ei ole pakollinen
”should”	Vahva ehdotus
”the following”	Listan aloitus

Perusmuodossaan RuleSpeak-sääntö muodostetaan seuraavasti:

<subjekti><avainsanat><kriteerit>(+<ehdot>)

Esimerkiksi ELF-projektin laatusääntö TRA03:ssa lentokenttään yhdistetyt solmut (AerodromeNode) pitää olla yhdistetty myös raitio- tai maanteihin. RuleSpeakilla sääntö määritellään seuraavasti:

TRA03

*Each AerodromeNode feature must intersect
the end point of at least one feature of one
of the following types:*

**RoadLink*

**RailwayLink*

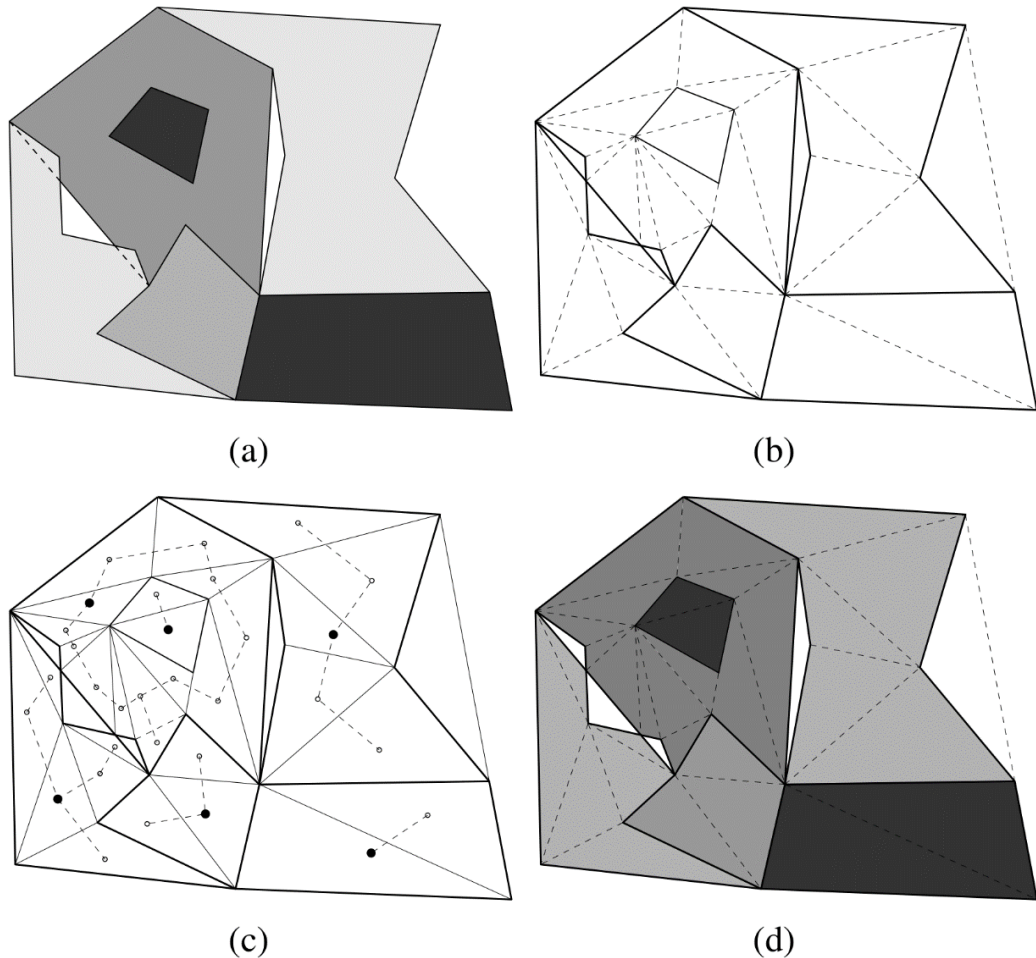
2.9 Työvälineitä laadun arviointiin

Paikkatietojen automaattinen laadunarviointi on konseptina melko uusi. Tästä syystä siihen tarkoitettuja valmiita ohjelmistoja ei ole kovin useita tarjolla. Kaupallisia, valmiisiin laatusääntöihin pohjautuvia ohjelmistoja ovat esimerkiksi Esri ArcGIS Data Reviewer sekä 1Spatial Cloud ja 1Validate. Täydellistä ja kaikki tapaukset kattavaa avoimen lähdekoodin ohjelmistoa ei ole tarjolla, mutta sellaisen tekemisen mahdollistavia ohjelmistoja on toki olemassa.

Esri ArcGIS Data Reviewer on suositun ArcGIS-paikkatieto-ohjelmiston laajennus, joka mahdollistaa laatutestien tekemisen graafisen käyttöliittymän ja muokattavissa olevien, valmiiden sääntöjen avulla. Laaditut testit voidaan tallentaa kerralla ajettavaksi koekelmäksi, jota kutsutaan Batch Jobiksi. Testien tuloksia voidaan tarkastella ArcMap-ohjelmassa. (ArcGIS Resources 2015.)

1Spatial Cloud ja **1Validate** ovat 1Spatialin paikkatiedon laatuun keskittyviä sovelluksia. Myös näissä laadunarviointi perustuu valmiisiin sääntöihin. Ajetuista testeistä saadaan laaturaportti sekä virheelliset kohteet sisältävä datasetti, jota voidaan tarkastella haluamallaan GIS-ohjelmistolla. 1Validate on ohjelmiston 1Spatial Management Suite -kokonaisuuteen liittyvä työpöytäversio, ja Cloud ESDIN-projektista lähtöisin oleva pilvipalvelu. (1Spatial 2015.)

Pprepair on avoimen lähdekoodin ohjelmisto, jolla voidaan tarkastaa ja korjata polygoniaineistojen topologiaa. Löydettäviä virheitä ovat virheelliset polygonit, raot, päällekkäisyydet sekä irrallaan olevat polygonit. Aineiston polygonien pitää siis muodostaa topologisesti yhtenäinen alue, jotta pprepair ei palauttaisi virheitä. Laskentamenetelmänä pprepair hyödyntää rajoitettua Delaunay-kolmiointia. Rajoitetun Delaunay-kolmioinnin periaate paikkatietojen virheiden etsinnässä esitetään kuvassa 2. Pprepair muodostaa polygonien kattamasta alueesta kolmioverkon, jossa jokaiselle kolmiolle annetaan polygonia vastaava tunnus Depth First (DFS) -haun perusteella. Virheinä ilmenevät sellaiset polygonit, joille ei ole annettu tunnusta. Mikäli kolmiointiin sisältyy kolmioita polygonien määräämän tason ulkopuolta, merkitään näiden kolmioiden tunnuksiksi, että ne kuuluvat ulkopuoleen. Siten ulkopuoli ei esiinny virheinä, vaikka se ei tasoon kuulukaan. (Arroyo Ohori et al. 2012.)



Kuva 2. Virheiden löytäminen rajoitetun Delaunay-kolmioinnin avulla. a) Polygonien muodostama pinta. b) Rajoitettu kolmiointi. c) Kolmioille määrätään tunnukset DFS-haulla. d) Kolmioinnin lopputulos, jossa virheellisillä, valkoisilla kolmioilla ei ole määrättyjä tunnuksia. (Ledoux & Meijers 2010.)

Automaattinen laadunarviointi voidaan toteuttaa myös avoimen lähdekoodin paikkatieto-ohjelmilla. JTS Topology Suite on paikkatietoanalyysiin keskittyvä ohjelmointikirjasto, joka on monissa avoimen lähdekoodin ohjelmistoissa hyvin keskeisessä roolissa. JTS kuuluu muun muassa OGC:n vapaiden GIS-työkalujen GeoTools-pakettiin sekä PostGIS-tietokantaohjelmistoon. JTS toteuttaa OGC:n Simple Features for SQL -standardin, joten sen avulla voidaan tehdä hyvin monipuolisia paikkatietoanalyyskejä. Tämän ansiosta se sopii hyvin alustaksi automaattiseen paikkatietojen laadunarviointiin. JTS:n historiaa on kerrottu osana lukua 4.2. (Vivid Solutions 2015b, OSGeo 2015.)

3 Spatiaaliset relaatiot

Tässä luvussa kerrotaan paikkatietoanalyysissä käytetyistä spatiaalisista relaatioista, jotka kuvaavat paikkatietokohteiden topologisia suhteita toisiinsa nähden. Topologisen eheyden testaus edellyttää ymmärrystä relaatioista.

3.1 Tausta

Ihminen hahmottaa luonnostaan, miten ympäristössään olevat asiat sijoittuvat toisiinsa nähden. Esimerkiksi voitaisiin havaita pöytä joka on maton päällä. Matto ja pöytä ovat molemmat huoneen keskellä, ja niiden päällä katosta roikkuu valaisin. Huoneeseen astuessaan ihminen tietää heti, että matto koskettaa lattiaa ja pöytä mattoa, ja että lamppu ei kosketa kumpaakaan. Kartalla ihminen näkee paikkatietokohteiden välisiä suhteita, esimerkiksi vierekkäin olevia kuntia ja järvien sisällä olevia saaria. Näitä asioiden välisiä, topologisia suhteita kutsutaan spatiaalisiksi relaatioiksi. Spatiaaliset relaatiot on kehitetty vastaamaan ihmisen käsitystä asioiden välisistä suhteista tilassa. Niiden avulla tehtävä paikkatietoanalyysi on alan keskeisimpiä asioita, ja myös äärimmäisen tärkeässä roolissa tässä työssä tehtävässä paikkatietojen laadunarvioinnissa. Loogisen eheyden validointi ja PostGIS:n ominaisuuksien riittävän hallinnan omaksuminen edellyttää relaatioiden tuntemusta ainakin perustasolla.

Spatiaalisia relaatioita on aikojen saatossa pyritty kuvaamaan erilaisin menetelmin, mutta tietokantaratkaisuissa vakiintuneet käytännöt ovat peräisin vasta 1990-luvun alusta, jolloin paikkatietojärjestelmien monimutkaistuminen aiheutti suurta painetta spatiaalisten relaatioiden muodollisen kuvaamisen kehittämiseen. Aiemmin relaatioiden kuvaamiseen ei ollut varsinaisesti mitään vakiintunutta käytäntöä. Tarpeeseen pyrittiin vastaamaan laajan tutkimustyön tuloksena kehitetyllä neljän leikkauksen 4IM- (4 Intersection Model), ja myöhemmin yhdeksän leikkauksen 9IM-malleilla (Egenhofer et al. 1993). Egenhoferin yhdeksän relaation malli havaittiin edelleen osin puutteelliseksi kuvaamaan kaikkia relaatioita, joten se laajennettiin kohteiden ulottuvuuksia huomioivaksi DE-9IM-malliksi (Dimensionally Extended nine-Intersection Model)(Clementini et al. 1993, Clementini & Di Felice 1995).

Egenhoferin ja Clementinin yhteisessä tutkimuksessa muodostettiin perusta mallien käytölle kyselykielissä (Clementini et al. 1994). Täysin vastaavaa mallia käytetään relaatioiden käsittelyyn nykyisin muun muassa PostGIS:ssä ja Oracle Spatialissa. PostGIS:n DE-9IM löysi tiensä OGC:n Simple Features -määrittelystä (OGC 06-104r4 2010).

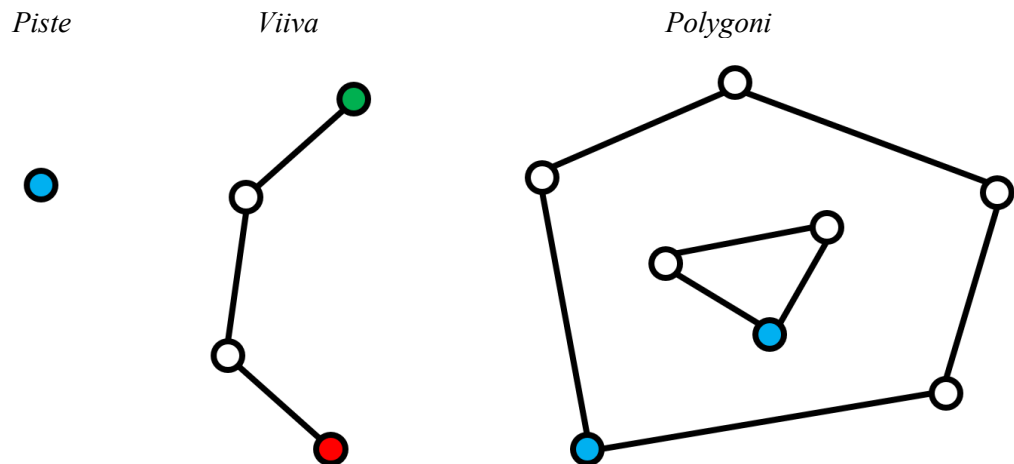
3.2 Paikkatietokohteiden geometriset kuvaukset

Geometrinen kuvaus on olennainen osa paikkatietoa, sillä paikkaan sidotun tiedon sijaintia ei sovelluksesta riippuen pystytä hyödyntämään erityisen tehokkaasti, jos kohteen muotoa ja asentoa ei tunneta. Vektorimuotoisten paikkatietokohteiden geometriset kuvaukset voidaan jakaa karkeasti kolmeen tyyppiin: Pisteisiin, viivoihin ja alueisiin. Paikkatietokohteiden geometriset kuvaukset esitetään kuvassa 3. Kaikkien geometrysten muotojen voidaan kuitenkin kuvitella koostuvan pisteistä. Lisäksi kohteet voivat olla joko yksittäisiä geometrisia muotoja tai useasta muodosta koostuvia multikohteita. Multikohdetta kutsutaan geometriakokoelmaksi (engl. geometry collection), jos sen geometria koostuu eri tyyppisistä kohteista. Clementinin ja Di Felicen (1995) mukaan paikkatietokohteiden geometriset kuvaukset määritellään spatiaalisissa relaatioissa seuraavasti.

Alueet ovat yksinkertaisesti yhdistettyjä sulkeumia kaksiulotteisessa avaruudessa \mathbb{R}^2 . Tässä työssä alueista käytetään myös termiä polygoni, sillä yksinkertaisena geometrisena muotona alue esitetään polygonina. Jos alueen geometrinen kuvaus sisältää useampia polygonoja, kyseessä on multipolygoni. Alueen reuna on suljettu käyrä, sisäpuoli rajan sisään jäävä ala ja ulkopuoli kaikki, joka ei kuulu näihin. Jokaisessa polygonin kulmassa on solmu. Aukkojen reunat ovat suljettuja käyriä alueen sisäpuolella. (Clementini & Di Felice 1995.)

Viivat ovat yksiulotteisia pistejoukkoja \mathbb{R}^2 :ssa. Viivat eivät leikkaa itseään ja niillä on tasan kaksi päätepistettä. Viivan rajana ovat vain sen päätepisteet ja sisäosana niiden väliin jäävä osa viivasta. Ulkopuolelle jäävät kaikki pisteet, jotka eivät kuulu sisäosan tai rajan pisteisiin. Useamman viivan sisältävää kohdetta kutsutaan multiviivaksi (engl. multilinestring). Jokaisessa viivan taitteessa sekä sen alku- ja loppupisteissä on solmut. (Clementini & Di Felice 1995.)

Piste on tasan yksi nollaulotteinen kohde \mathbb{R}^2 :ssa. Pisteellä ei ole reunaa lainkaan, joten sen geometria tulkitaan ainoastaan sisäosaksi. Ulkopuoleen kuuluu siis kaikki sisäosan pistettä lukuun ottamatta. Useamman pisteen sisältävää kohdetta kutsutaan multipisteeksi (engl. multipoint). (Clementini & Di Felice 1995.)



Kuva 3. Paikkatietokohteiden geometriset kuvaukset. Pisteen geometria on vain yksi piste. Viivan alkupiste on merkitty vihreällä ja loppupiste punaisella. Polygonin ja sen aukon reunojen lähtö- ja alkupisteet ovat samat.

3.3 Spatiaalisten relaatioiden esitystavat

Spatiaalisten relaatioiden esitystavat perustuvat pääosin pistejoukkotopologiaan. Relaatiot voidaan esittää predikaattifunktioina tai joukko-oppia mukailevina korkean tason matemaattisina ilmaisuina. Predikaatit ovat eräänlaisia esikonfiguroituja relaatiomatriiseja, jotka on pyritty nimeämään selkeästi, jotta tavallinenkin käyttäjä voisi käyttää niitä helpommin. Predikaattifunktioiden avulla kuvataan viittä erilaista relaatiota, joihin kuuluvat Equals-, Disjoint-, Touches-, Contains sekä Covers-relaatiot. Tapauksien määrää voidaan laajentaa matriisitranspoosin tai loogisen negation avulla kattamaan Intersects-, Within- sekä CoveredBy-relaatiot. Kohteiden ulottuvuudet huomioon ottaen voidaan predikaattien joukkoon lisätä vielä Crosses- ja Overlaps-relaatiot. (OGC 06-103r4

2011.) Kaikkiaan DE-9IM kattaa siis kymmenen erilaista tapausta. Matriisien avulla voidaan toki nimellisesti määritellä lukuisia relaatioita, mutta predikaattien kehittämisessä on huomioitu niiden tarkoituksenmukaisuus. Kaikki lukuisat relaatiot eivät ole käyttökelpoisia tai niillä on päällekkäisiä käyttötarkoituksia, mistä syystä niiden määrää on pyritty rajaamaan vain hyödyllisiin tapauksiin. (Clementini & Di Felice 1995.)

Taulukko 3. Relaatioiden nimet SFS ja SQL/MM -standardeissa sekä Egenhoferin tutkimuksessa.

Simple Features	Egenhofer	SQL/MM
Equals	Equal	ST_Equals
Disjoint	Disjoint	ST_Disjoint
Intersects	\neg Disjoint	ST_Intersects
Touches	Meet	ST_Touches
Within	Inside, CoveredBy	ST_Within
Contains	Contains, Covers	ST_Contains
Overlap	Overlap	ST_Overlaps

Taulukossa 2 on rinnakkain relaatioiden nimiä Simple Features -standardista (OGC 06-103r4 2011), Egenhoferin tutkimuksesta (Egenhofer et al. 1994) sekä SQL/MM -standardista (ISO/IEC 13249-3 2011). Relaatioiden nimet ja määrittelyt poikkeavat vain vähän toisistaan, joten voidaan todeta, että määrittelyiden perusta on pysynyt samana tähän päivään asti.

Spatiaalisia relaatioita voidaan tarkastella leikkausmatriiseina. Clementini-, eli DE-9IM-matriisilla tarkoitetaan 3x3 matriisia, jonka solut kuuluvat arvoalueeseen {0, 1, 2, F}. Yksinkertaisempaa Egenhofer-, eli 9IM-matriisia voidaan pitää binäärisenä tai totuusarvoisena vastineena DE-9IM-matriisille. 9IM-matriisin arvot kuuluvat siten alueeseen {1, 0} tai {T, F}. Yksinkertaisempaa matriisia voidaan käyttää esimerkiksi optimointitarkoituksessa, mikäli syötteelle asetetaan tarkoituksenmukaisia rajoitteita. Lisäksi sellaisia soluja, joiden arvot eivät ole relaation kannalta oleellisia, merkitään yleensä *- , eli tähtisymbolilla. Clementinin, Sharman ja Egenhoferin (1994) tutkimuksessa tälle käytettiin myös kreikkalaista pikku-deltaa, eli δ -symbolia, mutta tietojärjestelmien tapauksessa kreikkalaisten kirjaimien käyttö saattaa aiheuttaa ongelmia. (Clementini & Di Felice 1995, Clementini et al. 1993.) Sekä Clementini- että Egenhofer-matriiseja käytetään paikatietojärjestelmissä laajasti.

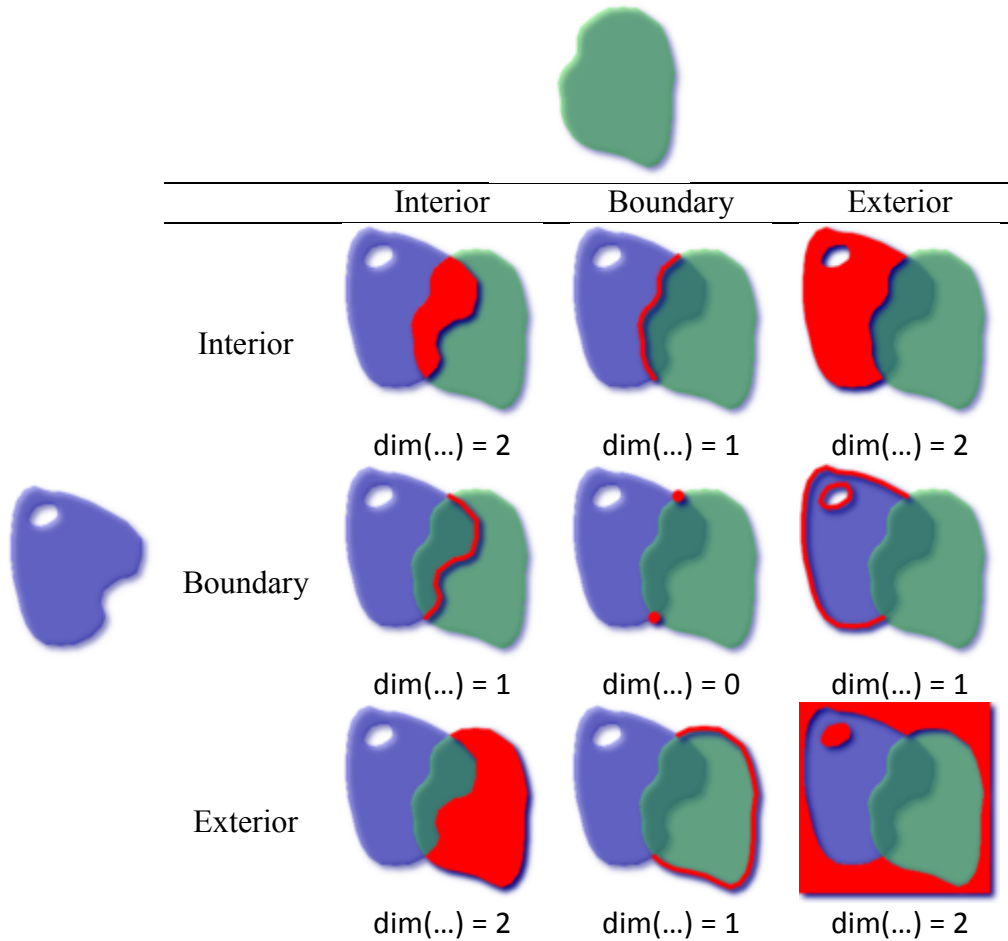
3.3.1 Leikkausmatriisien rakenne

Relaatioita määrittelevissä leikkausmatriiseissa solut edustavat kahden kohteen välisiä leikkauksia. Matriisissa vaakarivit kuvaavat kohdetta A ja sarakkeet kohdetta B siten, että ensimmäinen rivi esittää A:n sisäosaa, toinen rivi reunaa ja viimeinen ulkopuolta. Sarakkeet on aseteltu samoin vasemmalta oikealle. I, B ja E ovat kohteiden sisäosa (engl. interior), reuna (engl. boundary) sekä ulkopuoli (engl. exterior). Simple Features -standardissa matriisiin S merkittävät leikkauksen arvot L voivat kuulua kaavan 1 mukaisesti joukkoon {T, F, *, 0, 1, 2}. Varsinainen leikkausmatriisi muodostetaan kaavan 2 mukaisesti, jossa kohteiden osien tunnuskirjaimista I, B ja E muodostettu yhdiste määrittää relaation. (OGC 06-103r4 2011.)

$$L = \begin{cases} T & , \text{jos } \dim(S) \in \{0, 1, 2\} \\ F & , \text{jos } \dim(S) = -1, \text{ eli kun } S = \emptyset \\ * & , \text{jos } \dim(S) \in \{-1, 0, 1, 2\}, \text{ eli mikä tahansa} \\ 0 & , \text{jos } \dim(S) = 0, S \text{ sisältää pisteen, mutta ei viivoja} \\ 1 & , \text{jos } \dim(S) = 1, S \text{ sis. ainakin viivan, mutta ei alueita} \\ 2 & , \text{jos } \dim(S) = 2, S \text{ sis. ainakin alueen} \end{cases} \quad (1)$$

$$RM = \begin{bmatrix} II & IB & IE \\ BI & BB & BE \\ EI & EB & EE \end{bmatrix} \quad (2)$$

Kuten luvussa 3.1 kerrotaan, pisteen dimensio on 0, viivan 1 ja alueen 2. Leikkausmatriisissa tuloksen dimensio perustuu kohteiden leikkausten tarkasteluun. Kuvassa 4 esitetään esimerkkinä kahden polygonin leikkausrelaatio, jossa DE-9IM-predikaattimaskiksi muodostuu '212101212', kun ulottuvuudet otetaan huomioon (The PostGIS Development Group 2015).



Kuva 4. DE-9IM-leikkaukset kuvina kahden polygonin leikatessa (The PostGIS Development Group 2015).

$$DE9IM(a, b) = \begin{bmatrix} \dim(a^0 \cap b^0) & \dim(a^0 \cap \partial b) & \dim(a^0 \cap b^e) \\ \dim(\partial a \cap b^0) & \dim(\partial a \cap \partial b) & \dim(\partial a \cap b^e) \\ \dim(a^e \cap b^0) & \dim(a^e \cap \partial b) & \dim(a^e \cap b^e) \end{bmatrix} \quad (3)$$

$$\begin{aligned} \text{bin}(DE9IM(a, b)) = 9IM(a, b) = \\ \begin{bmatrix} a^0 \cap b^0 \neq \emptyset & a^0 \cap \partial b^0 \neq \emptyset & a^0 \cap b^e \neq \emptyset \\ \partial a^0 \cap b^0 \neq \emptyset & \partial a^0 \cap \partial b^0 \neq \emptyset & \partial a^0 \cap b^e \neq \emptyset \\ a^e \cap b^0 \neq \emptyset & a^e \cap \partial b^0 \neq \emptyset & a^e \cap b^e \neq \emptyset \end{bmatrix} \end{aligned} \quad (4)$$

Kaavassa 3 esitetään, miten varsinainen DE-9IM-leikkausmatriisi muodostetaan. Jokaisessa matriisin solussa on leikkauksen dimensio. Binäärimuotoinen leikkausmatriisi, eli Egenhofer-matriisi saadaan yksinkertaisesti muuttamalla Clementini-matriisin dimensiot binäärisiksi kaavan 4 mukaisesti. Tällöin soluihin merkitään T, eli True, jos leikkaus on olemassa, ja False, mikäli se ei ole olemassa. Jos relaatiolla ei ole merkitystä tarkasteltavan asian kannalta, merkitään soluun tähtisymboli sekä Egenhofer- että Clementini-matriisien yhteydessä. (Clementini & Di Felice 1995, The PostGIS Development Group 2015.)

3.3.2 Predikaatit joukossa {T, F, *}

Equals- eli yhtäsuuruuspredikaatti kertoo, ovatko tarkasteltavien kohteiden geometriset kuvaukset identtisiä. Kohteiden A ja B geometriat ovat identtiset, kun niiden sisäosat leikkaavat, ja ulko-osat voivat leikata vain keskenään. Equals-predikaatin matemaattinen esitys ja leikkausmatriisi esitetään kaavan 5 mukaisesti. Tällöin relaation maskiksi muodostuu 'T*F***FFF*'. Equals-predikaatti vastaa Within- ja Contains-predikaattien yhdistelmää. (Vivid Solutions 2015a.)

$$a. Equals(b) \Leftrightarrow (a \cap b = a) \wedge (a \cap b = b) \Leftrightarrow \begin{bmatrix} T & * & F \\ * & * & F \\ F & F & * \end{bmatrix} \quad (5)$$

Disjoint- eli erillisuuspredikaatti kertoo, ovatko tarkasteltavat kohteet erillään toisistaan. Kohteet ovat erillään, jos vain niiden ulko-osat voivat leikata keskenään. Disjoint-predikaatin matemaattinen esitys ja leikkausmatriisi esitetään kaavan 6 mukaisesti. Relaation maskina on tällöin 'FF*FF*****'. (Clementini et al. 1993, OGC 06-103r4 2011.) Disjoint-predikaatin käänteispredikaatti on Intersects.

$$a. Disjoint(b) \Leftrightarrow (a \cap b = \emptyset) \Leftrightarrow \begin{bmatrix} F & F & * \\ F & F & * \\ * & * & * \end{bmatrix} \quad (6)$$

Touches- tai meets, eli vierekäisyuspredikaatti kertoo, koskettavatko kohteet toisiaan. Kohteet koskettavat, jos niillä on ainakin yksi sama piste, mutta niiden sisäosat eivät leikkaa. Sisäosat eivät saa leikata tässä relaatiossa, joten piste ja toinen piste eivät voi koskettaa toisiaan, sillä pisteellä ei ole reunoja lainkaan. Touches-relaatiolla on kolme mahdollista tapausta, joista ensimmäisessä kohteen A sisäosa leikkaa kohteen B reunaa. Toisessa kohteen A reuna leikkaa kohteen B sisäosaa. Kolmannessa kohteiden A ja B reunat leikkaavat keskenään. Matemaattisesti tämä voidaan esittää kaavan 8 mukaisesti, jolloin relaatioiden maskeiksi muodostuvat 'FT*****', 'F**T*****' sekä 'F***T*****'. (Clementini et al. 1993, OGC 06-103r4 2011.)

$$\begin{aligned} a. Touches(b) \Leftrightarrow (a \cap b \neq \emptyset) \wedge (a^0 \cap b^0 = \emptyset) \Leftrightarrow \\ \begin{bmatrix} F & T & * \\ * & * & * \\ * & * & * \end{bmatrix}, \begin{bmatrix} F & * & * \\ T & * & * \\ * & * & * \end{bmatrix}, \begin{bmatrix} F & * & * \\ * & T & * \\ * & * & * \end{bmatrix} \end{aligned} \quad (7)$$

Within- tai Inside-predikaatin avulla tarkastellaan, onko kohde A kohteen B sisällä. Jos A:n todetaan olevan B:n sisällä, ja myös B:n A:n sisällä, ovat kohteet geometrialtaan samat. Within-predikaatin matemaattinen esitys on kaavan 8 mukainen, ja sitä vastaava predikaattimaski 'T**F**F***'. (Clementini & Di Felice 1995, Vivid Solutions 2015a.)

$$a. Within(b) \Leftrightarrow (a \cap b = a) \wedge (a^0 \cap b^0 \neq \emptyset) \Leftrightarrow \begin{bmatrix} T & * & F \\ * & * & F \\ * & * & * \end{bmatrix} \quad (8)$$

Covers-predikaatti on käytännössä toiminnaltaan lähes sama kuin Contains-predikaatti. Covers ei erottele reunojen tai sisäosan pisteitä Contains-predikaatin tavoin, vaan tutkii ainoastaan, onko kohde B kokonaan kohde A:n sisällä. Covers-predikaatin matemaattinen esitys on kaavan 9 mukainen, jolloin sen relaatioiden maskit ovat 'T*****FF*', '*T*****FF*', '***T**FF*' tai '*****T*FF*'. (Egenhofer et al. 1993, Vivid Solutions 2015a.)

$$a. Covers(b) \Leftrightarrow (a \cap b = b) \Leftrightarrow \begin{bmatrix} T & * & * \\ * & * & * \\ F & F & * \end{bmatrix}, \begin{bmatrix} * & T & * \\ * & * & * \\ F & F & * \end{bmatrix}, \begin{bmatrix} * & * & * \\ T & * & * \\ F & F & * \end{bmatrix}, \begin{bmatrix} * & * & * \\ * & T & * \\ F & F & * \end{bmatrix} \quad (9)$$

3.3.3 Käänteispredikaatit

Intersects-, eli leikkauspredikaatti kertoo, leikkaavatko kohteet keskenään millään tavalla. Intersects-predikaatti kattaa Overlaps-, Touches- ja Within -tapaukset, eli Intersects pätee kohteiden ollessa limittäin, koskettaessa toisiaan tai kun ne ovat sisäkkäin. Intersects voidaan edelleen jakaa neljään eri tapaukseen, joita merkitään matriiseina ja lauseina eri tavoin. Leikkauksen oletetaan tapahtuvan, kun kohteiden sisäosat, kohteen A sisäosa ja kohteen B reuna, kohteen A reuna sekä B:n sisäosa tai kohteiden A ja B reunat leikkaavat. Vastaavat matemaattiset esitykset ja leikkausmatriisit esitetään kaavan 10 mukaisesti, jolloin relation maskit voidaan merkitä 'T*****', '*T*****', '***T*****' tai '*****T*****'. Intersects on käänteinen Disjoint-predikaatti, eli tapaus voidaan tulkita myös siten, että kohteet eivät ole erillään, eli kun niiden leikkaus ei ole tyhjä joukko. (The PostGIS Development Group 2015, Vivid Solutions 2015a.)

$$a. Intersects(b) \Leftrightarrow (a \cap b \neq \emptyset) \Leftrightarrow \begin{bmatrix} T & * & * \\ * & * & * \\ * & * & * \end{bmatrix}, \begin{bmatrix} * & T & * \\ * & * & * \\ * & * & * \end{bmatrix}, \begin{bmatrix} * & * & * \\ T & * & * \\ * & * & * \end{bmatrix}, \begin{bmatrix} * & * & * \\ * & T & * \\ * & * & * \end{bmatrix} \quad (10)$$

Contains- eli sisältyvyyspredikaatti kertoo, onko kohde B kohde A:n sisällä. Relaatiassa kohteiden sisäosat leikkaavat keskenään, mutta B:n sisäosa tai raja eivät saa leikata A:n ulko-osan kanssa. Contains voidaan tulkita käänteiseksi Within-predikaatiksi, jossa kohteiden sisältyvyysjärjestys vaihtuu, eli tällöin B sisältää A:n. Contains voidaan ilmaista kaavan 11 mukaisesti, jolloin sen predikaattimaski on 'T*****FF*'. (OGC 06-103r4 2011.)

$$a. Contains(b) \Leftrightarrow b. Within(a) \Leftrightarrow (a \cap b = b) \wedge (a^0 \cap b^0 \neq \emptyset) \Leftrightarrow \begin{bmatrix} T & * & * \\ * & * & * \\ F & F & * \end{bmatrix} \quad (11)$$

CoveredBy-predikaatti on Covered-predikaatin käänteispredikaatti, joka pätee kohteiden A ja B leikkauksen ollessa sama kuin A:n geometria. CoveredBy voidaan ilmaista kaavan 12 mukaisesti, jolloin sen predikaattimaskiksi muodostuu 'T*F**F****', '*TF**F****', '**FT*F****' tai '**F*TF****'. (Egenhofer et al. 1993, Vivid Solutions 2015a.)

$$a.CoveredBy(b) \Leftrightarrow b.Covers(a) \Leftrightarrow (a \cap b = a) \Leftrightarrow$$

$$\begin{bmatrix} T & * & F \\ * & * & F \\ * & * & * \end{bmatrix}, \begin{bmatrix} * & T & F \\ * & * & F \\ * & * & * \end{bmatrix}, \begin{bmatrix} * & * & F \\ T & * & F \\ * & * & * \end{bmatrix}, \begin{bmatrix} * & * & F \\ * & T & F \\ * & * & * \end{bmatrix} \quad (12)$$

3.3.4 Predikaatit joukossa {0, 1, T, F, *}

Crosses-, eli risteävyyspredikaatti kertoo risteävätkö kohteiden geometriat keskenään. Kohteet risteävät, jos kohteiden sisäpuolet leikkaavat keskenään ja toisen kohteen sisäpuoli leikkaa toisen kohteen sisäpuolta. Kohteet risteävät myös, jos niiden sisäpuolten leikkaus on 0-ulotteinen, eli piste. Kyseessä on tällöin kahden viivan välinen risteäminen. Crosses ilmaistaan kaavan 13 mukaisesti, ja sen predikaattimaskit ovat 'T*T*****', 'T*****T**' ja '0*****'. (ISO/IEC 13249-3. 2011.)

$$a.Crosses(b) \Leftrightarrow (\dim(a^0 \cap b^0) < \max(\dim(a^0), \dim(b^0))) \wedge$$

$$(a \cap b \neq a) \wedge (a \cap b \neq b) \Leftrightarrow$$

$$\begin{bmatrix} T & * & T \\ * & * & * \\ * & * & * \end{bmatrix}, \begin{bmatrix} T & * & * \\ * & * & * \\ T & * & * \end{bmatrix}, \begin{bmatrix} 0 & * & * \\ * & * & * \\ * & * & * \end{bmatrix} \quad (13)$$

Overlaps- eli limittäisyyspredikaatti vertaa kahta kohdetta samassa ulottuvuudessa. Jos kohteiden erotuksen geometria on erilainen kuin kumpikaan kohde, ovat kohteiden geometriat limittäin. Relaatiossa kohteiden A ja B sisäosat leikkaavat keskenään sekä toisensa ulko-osien kanssa. Matemaattisesti tämä voidaan ilmaista kaavan 14 mukaisesti. Relaation maskiksi muodostuu tällöin 'T*T***T**'. (OGC 06-103r4 2011.)

$$a.Overlaps(b) \Leftrightarrow (a \cap b \neq a) \wedge (a \cap b \neq b) \Leftrightarrow \begin{bmatrix} T & * & T \\ * & * & * \\ T & * & * \end{bmatrix} \quad (14)$$

4 PostGIS

PostGIS on paikkatietolaajennus PostgreSQL-tietokannalle, joka mahdollistaa sen käyttämisen paikkatietokantana kartta- ja paikkatietosovelluksille. Tässä luvussa perehdytään PostGIS:n kehitykseen nykymuotoonsa asti ja kerrotaan ohjelmiston tärkeimmistä ominaisuuksista. Laadunarvioinnin kannalta PostGIS:n olennaisin ominaisuus on sen spatiaalisiin relaatioihin perustuvat spatiaaliset funktiot, jotka esitellään luvussa kattavasti. Näitä funktioita on käytetty erittäin laajasti työn käytännön osuudessa.

4.1 Spatiaaliset tietokannat ja PostGIS

Spatiaalinen tietokanta tai paikkatietokanta tarkoittaa sellaista tietokantaa, johon voidaan tallentaa geometristä dataa spatiaalisten tietotyyppien muodossa. Tyypillisesti spatiaaliset tietokannat ovat relaatiotietokantoja, jotka tarjoavat erikoistuneita funktioita ja indeksejä paikkatietojen kanssa työskentelyyn. (Obe & Hsu 2011, s. 5.) Paikkatieto-ominaisuuksilla varustettuja tietokantoja on lukuisia, ja näitä ovat muun muassa Spatialite, Oracle Spatial sekä PostGIS.

Spatiaalinen tietokanta on toki sähköinen tietovaranto, kuten kaikki muutkin tietokannat ovat. Erot eivät jää vain siihen, että spatiaalisuus tarkoittaa, että tietokantaan voidaan tallentaa myös paikkatietokohteita. Suurimpia käytännön eroja ovat erot kyselyiden monipuolisuudessa, sillä tavanomaisella tietokannalla ei ole spatiaalisen tietokannan kykyä paikkatietoanalyysiin. Monipuolisten, kansainvälisiin standardeihin perustuvien spatiaalisten funktioiden ja operaattorien ansiosta PostGIS-tietokannan analyysimahdollisuudet vastaavat modernia paikkatieto-ohjelmaa. (Ramsey 2007.)

PostGIS on GNU GPL -lisenssiä käyttävä avoimen lähdekoodin ohjelmisto, joka tarjoaa PostgreSQL-relaatiotietokantaohjelmistolle tuen paikkatiedoille. PostGIS toteuttaa Open Geospatial Consortiumin (OGC) Simple Features for SQL (SFS) sekä ISO/IEC 13249-3:2011 SQL/MM Part 3: Spatial -standardit. OGC on kansainvälinen, tuottoa tavoittelematon standardointiorganisaatio, joka on keskittynyt paikkatietoalan avoimiin standardeihin. Avoimet standardit ovat vapaasti kaikkien käytettävissä, sillä niiden tarkoitus on parantaa paikkatietojen yhteensopivuutta maailmanlaajuisesti. PostGIS:n keskeisiä ominaisuuksia ovat geometriatyytit pisteille, viivoille, polygoneille sekä näiden moniversiolle ja geometriakokoelmille, spatiaaliset predikaatit ja operaattorit analyysiä varten, spatiaalinen indeksointi kyselyiden nopeuttamiseksi sekä ominaisuudet rasteridatan hallintaan. (Ramsey 2007.)

Relaatiotietokannalla tarkoitetaan sellaista tietokantaa, joka käyttää predikaattilogiikkaan perustuvaa relaatiomallia. Tietokannassa tiedot on tallennettu taulumuodossa, jossa sarakkeet edustavat arvoja ja rivit tallennusyksiköitä. Relaatiotietokannassa taulut yksilöidään, jotta niihin voitaisiin viitata, ja että viittauksen avulla eri taulujen tietoja voitaisiin käyttää myös yhdessä. Lähes kaikki relaatiotietokannat tukevat SQL-kyselyitä jossain muodossa sekä tiedon noutamista että tietokannan hallintaa varten. (Garcia-Molina et al. 2009, s. 4.)

PostGIS-kyselyiden suorittamiseen tarvitaan samat peruselementit menetelmästä riippumatta. Itse kyselyt voidaan toki kirjoittaa esimerkiksi millä tahansa tekstieditorilla, tallentaen valmiin kyselyn teksti- tai SQL-tiedostona. Kysely välitetään PostGIS-palvelimelle komentorivin tai graafisen ohjelman kautta, mutta se voidaan sisällyttää myös ohjelmakoodiin, jos kysely on tarkoitus suorittaa esimerkiksi Javalla. (Ramsey 2007.)

PostGIS:n peruskäyttö on työssä kuvatuilla menetelmillä vaivatonta. PostGIS:n käyttöönotto edellyttää PostgreSQL- ja PostGIS-asennukset. PostGIS:n lataussivustolla on saatavilla yksityiskohtaiset ohjeet näiden lataamiseen ja asentamiseen eri ympäristöissä. Tässä työssä PostGIS asennettiin Windows-ympäristölle, mutta sitä on saatavilla myös OSX- ja Linux paketteina sekä itse käännettävänä lähdekoodina (PostGIS 2015a).

4.2 PostGIS kehittyi

PostGIS:n kehittäminen lähti tarpeesta lisätä paikkatieto-ominaisuuksia laajentamaan PostgreSQL:n varsin rajoittuneita, paremmin tietokonegrafiikkaan soveltuvia geometrisiä tyyppejä. Refrations Research aloitti kehitystyön vuonna 2000 Brittiläisen Kolumbian hallinnolle SFS-standardin pohjalta. (Refrations Research 2015.)

Uusien tyyppien lisääminen on PostgreSQL:n keskeisiä ominaisuuksia. Lisäksi se sisältää GiST-indeksointimallin, joka sopi erityisen hyvin spatiaalisen indeksin mahdollistamiseen. Näiden komponenttien avulla pystyttiin luomaan perusta PostGIS:lle, jonka ensimmäinen versio 0.1 lopulta julkistettiin 31.5.2001. Ensimmäinen PostGIS sisälsi SFS-standardiin pohjautuvan Well Known Text (WKT) -esityksen paikkatietokohteille, spatiaalisen indeksin nopeampaa aineiston noutamista varten, joitain perusfunktioita datan käsittelyyn sekä yksinkertaisen Java-laajennuksen, jonka avulla voidaan hallita tietokantaa. (Refrations Research 2015.)

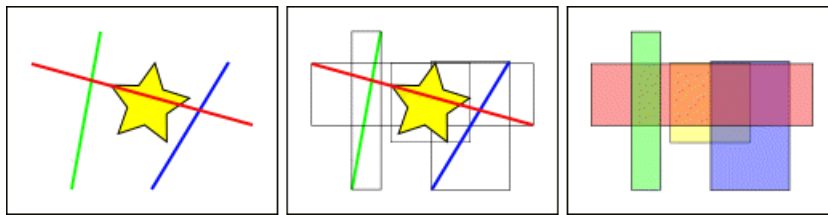
Vaikeimmin toteutettavat funktiot, eli spatiaaliset predikaatit ja operaattorit, kehitettiin omana rinnakkaisprojektina myös Brittiläisen Kolumbian toimeksiannosta. Vivid Solutions aloitti JTS Topology Suitein kehityksen suunnilleen samanaikaisesti PostGIS:n kanssa, minkä oli määrä toteuttaa samoja OGC:n standardeja kuin PostGIS. Lopulta vuoden 2001 lopulla työn tuloksena oli kattava, SFS:n mukainen kirjasto laskennallista geometriaa käyttäviä algoritmeja. Jotta alun perin Java-ohjelmointikielellä tehtyä JTS:ää voitiin käyttää PostGIS:ssä, jouduttiin siitä vielä kehittämään C++-kielinen versio Refrations Researchin ja Vivid Solutionsin yhteistyönä. Valmis C++-toteutus sai lopulta nimen Geometry Engine Open Source (GEOS). (Refrations Research 2015.)

Kuuden kehitysversion ja mittavan tietorakenteiden optimoinnin jälkeen PostGIS:n ensimmäinen julkaisuversio 1.0 saatiin yleisesti tarjolle vuonna 2005. Ensimmäisessä julkaisuversiossa oli nyt täysi tuki SFS-standardille GEOS:n mahdollistamilla funktioilla. Versioon 1.3 asti kehitys tähtäsi SFS-standardin sijaan tukemaan ISO SQL/MM -standardia, minkä ansiosta PostGIS päätyi tukemaan entistä useampia paikkatietokohteita, kuten kaarevia kohteita ja geografiatietotyyppiä, joka kuvaa kohteita maan pinnan suhteen. Paikkatietofunktioiden nimet muutettiin myös vastaamaan uutta määritelmää ST_-etuliitteineen. (Refrations Research 2015.)

PostGIS:n tehokkuutta on parannettu jatkuvasti erityisesti indeksien käytön osalta. PostGIS 2.0:sta lähtien pystytään muun muassa etsimään kohteita indeksin avulla lähimmän naapurin menetelmällä kolmi- ja neliulotteisesti, käyttämään rasteritietotyyppiä sekä käyttämään analyysiä vektori- ja rasteridatan välillä. (Refrations Research 2015.)

4.3 Spatiaalinen indeksi ja GiST

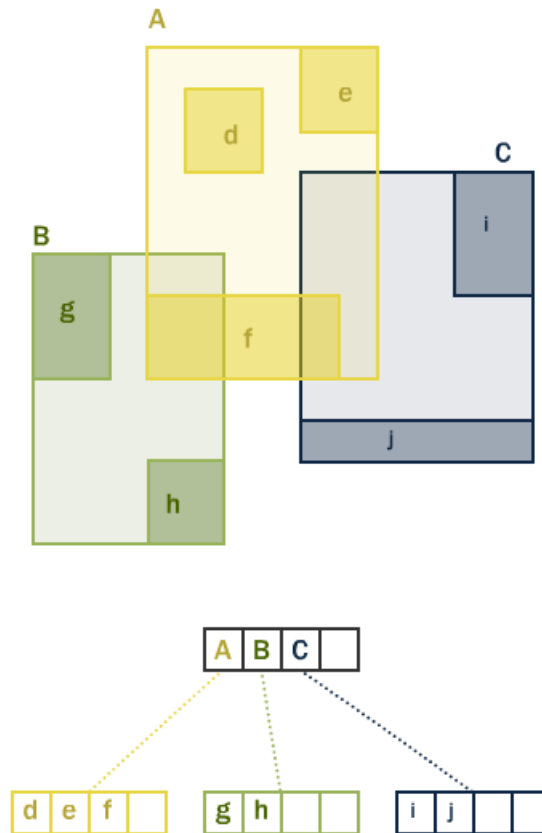
Indeksoinnin tarkoituksena on nopeuttaa hakuja tietokannoissa. Ilman indeksointia kyselyitä suorittaessa jouduttaisiin usein koko tietokanta käymään läpi lukuisia kertoja, joten ilman indeksointia suurten datamäärien prosessointiin ja säilytykseen tarkoitettut paikkatietokannat eivät olisi erityisen käyttökelpoisia. Ilman spatiaalista indeksiä kevyestäkin kyselystä saattaisi tulla erittäin raskas operaatio, jos jokaisen kohteen kohdalla käytäisiin tietokantaa läpi alusta alkaen, ja koko kyselyn aikana tuhansien hakujen sijaan jouduttaisiin suorittamaan yhteensä jopa useita miljardeja hakuja. (The PostGIS Development Group 2015, 4.6 Building Indexes.)



Kuva 7. Minimirajoitelaatikat (OpenGeo Suite 4.7 2015).

Tietokannoissa indeksien tietorakenteet on yleisimmin toteutettu B-tai R-puina. B-puut ovat hyödyllisiä, kun tiedot voidaan lajitella rationaalisesti, mutta samalla tästä syystä B-puut eivät sovellu monimutkaisille paikkatiedoille erityisen hyvin. R-puut jakavat tiedot kerroksittain nelikulmioiksi, ja edelleen nelikulmiot nelikulmioiksi ja niin edelleen. R-puun neliöiden jako perustuu kohteiden sijaintia rajaavien suorakaiteiden, tai rajoitelaatikoiden (engl. bounding box) käyttöön, joiden muodostus esitetään kuvassa 7. Yksittäinen rajoitelaatikko muodostetaan kohteen ympärille siten, että kohde rajoittuu täsmälleen suorakaiteen sisälle. Indeksä muodostetaan ryhmittelemällä yksittäisiä rajoitelaatikoita sijaintiensa perusteella suurempia kokonaisuuksia kattaviksi rajoitelaatikoiksi ja vielä niiden rajoitelaatikoiksi. Ryhmittelyn perusteella muodostetaan lopulta puurakenne kuvan 8 mukaisesti siten, että jokainen indeksin yksikkö viittaa suorakaiteen muotoiseen alaan tasolla. (The PostGIS Development Group 2015.)

R-tree Hierarchy



Kuva 8. R-puu muodostuu rajoitelaatikoista (OpenGeo Suite 4.7 2015).

GiST, eli yleistetty hakupuu (engl. Generalized Search Tree), ryhmittelee tiedot suhteessa toisiinsa. Ryhmittely tapahtuu relaatioalgebran sääntöjen tyyliä esimerkiksi asioihin jotka ovat jonkin sisällä, sisältävät jonkin asian tai leikkaavat keskenään. Puut voidaan usein määrittellä mille tahansa tietotyyppille, mutta GiST antaa vapauksia jopa kyselyiden laajentamiseen. (PostGIS 2015b, Hellerstein et al. 1995.) PostGIS:n indeksi yhdistää GiST:n R-puuhun luodakseen entistä vakaamman tietorakenteen spatiaalista indeksointia varten. PostgreSQL tarjoaa muitakin puurakenteita, mutta GiST:n joustavuus tekee siitä parhaan vaihtoehdon tähän tarkoitukseen. (The PostGIS Development Group 2015.)

Spatiaalinen indeksi voidaan ottaa käyttöön seuraavasti. INDEXNAME on käyttäjän valitsema nimi indeksille. TABLENAME on tietotaulun nimi, jolle indeksi luodaan. GEOMETRYCOLUMN on tietotaulun sarake, joka sisältää kohteiden geometriatiedot. (PostGIS 2015b.)

```
CREATE INDEX <INDEXNAME> ON <TABLENAME> USING GIST
(<GEOMETRYCOLUMN>);
```

PostGIS ei käytä jokaiselle funktiolle ja operaattorille automaattisesti indeksiä. Relatiofunktiot tyypillisesti käyttävät, mutta funktiota käytettäessä on syytä tarkistaa asia PostGIS:n käyttöohjeista. Kysely voidaan asettaa lisäehdolla käyttämään indeksiä &&-operaattorin avulla. (The PostGIS Development Group 2015.)

4.4 Kyselyiden rakenne

PostgreSQL-kyselyissä käytetään relaatiotietokantojen hallintaan tarkoitettua SQL-ohjelmointikieltä (Structured Query Language). Kyselyiden perustana voidaan pitää SELECT-FROM-WHERE-rakennetta, jossa SELECT ilmoittaa valinnan kohteen, FROM mistä tietotaulusta valinta tehdään ja WHERE millä ehdoilla valinta tehdään. Valinnan kohteena voi olla yksittäiset sarakkeet tai kokonainen tietotaulun rivi tai yksikkö, jolloin valinta ilmoitetaan tähtisymbolilla *. Tulokset voidaan järjestää ORDER BY -lauseella. (PostgreSQL 2015, 4.1 Lexical Structure.) Perusmuodossaan kyselyt noudattavat seuraavaa syntaksia.

```
SELECT * FROM B WHERE valintaehto
```

Tekstiä voidaan valita seuraavalla tavalla.

```
SELECT GN.*  
FROM "GeographicalNames"."NamedPlace" GN  
WHERE text LIKE '%Kaunissaari%';
```

Tässä kyselyssä haetaan paikannimiaineistosta GeographicalNames kaikki paikannimet NamedPlace, joiden nimeen text sisältyy Kaunissaari. Tuloksena halutaan koko tietotaulun rivi, eli kaikki tiedot kohteista. WHERE-lauseen ilmaisulla tarkoitetaan, että kysely koskee sellaisia rivejä, joiden text-nimisten sarakkeiden arvo on kuin (engl. like) Kaunissaari. Prosenttimerkki ilmaisee, että halutun ilmaisun molemmin puolin voi olla mitä tahansa tekstiä. Jos halutaan vain sellaiset rivit, joiden nimessä ei esiinny muuta kuin Kaunissaari, olisi WHERE-ehto muotoa text = 'Kaunissaari'. Jos haluttaisiin laskea kaikki Kaunissaari-nimiset instanssit, voidaan siihen käyttää COUNT()-funktioita. Jos halutaankin laskea Kaunissaari-nimisten instanssien pinta-alat yhteen, voidaan siihen käyttää SUM()-funktioita. PostgreSQL sisältää myös muita matemaattisia funktioita ja nimettyjä vakioita, kuten tuloksen pyöristävä ROUND() tai piin arvo Pi(). (PostgreSQL 2015, 4.2 Value Expressions.)

```
SELECT COUNT (*) ...
```

```
SELECT ROUND(SUM(ST_Area(*.geom))) ...
```

Kysely saattaa joskus palauttaa saman kohteen useita kertoja, varsinkin jos kysely on moniosainen. Jos kohde halutaan valita lopulliseen tulokseen tai kyselyn osassa vain yhden kerran, käytetään avainsanaa DISTINCT ennen kohteen määrittelyä. (PostgreSQL 2015, 6.1 Reference - SELECT.)

```
SELECT DISTINCT (*) ...
```

COALESCE-avainsanan avulla saadaan tietotaulusta ensimmäinen instanssi, joka on olemassa, eli jonka arvo ei ole null. Tästä on hyötyä esimerkiksi selvitetessä, onko taulu tyhjä, vai onko siihen tallennettu jotain. (PostgreSQL 2015, 9.13 Conditional Expressions.)

WITH-rakenne auttaa erityisesti selkeyttämään monimutkaisempia kyselyitä. Rakenteen avulla voidaan luoda SELECT-rakenteen ulkopuolisia, vain yhdessä kyselyssä voimassa olevia väliaikaisia tietotauluja, joihin voidaan viitata varsinaisessa kyselyssä. Rakenteen sisällä suoritetaan oma SELECT-kyselynsä. WITH-rakenteen tyyliä, väliaikaisia tauluja

kutsutaan yleisesti Common Table Expressioneiksi (CTE). WITH-kyselyitä voi olla useita peräkkäin pilkulla eroteltuna. Ensimmäinen kysely alkaa WITH-sanalla, jonka jälkeen valitaan tilapäiselle taululle nimi. AS-sanan jälkeen sulkujen sisälle muodostetaan taulun luova kysely. Seuraavia samaan rakenteeseen muodostettavia väliaikaisia tauluja varten ei merkitä WITH-sanaa lainkaan. (PostgreSQL 2015, 7.8 WITH Queries (Common Table Expressions).) Hyvin yksinkertaisena esimerkkinä voidaan kuvitella, että halutaan laskea ihmisten lukumäärä yhteen kahdesta eri kunnasta. Tehdään molemmista kunnista väliaikainen taulu ja valitaan summaa varten halutut sarakkeet.

```
WITH kunta1 AS (
    SELECT *
    FROM väestö
    WHERE id = 1
),
kunta2 AS (
    SELECT *
    FROM väestö
    WHERE id = 2
)
SELECT SUM(kunta1.ihmiset, kunta2.ihmiset)
FROM kunta1, kunta2;
```

Kyselyiden tuloksia voidaan liittää toisiinsa UNION-, INTERSECT- ja EXCEPT-avainsanoilla. Jotta tämä onnistuisi, tulee kyselyiden koskea samoja tietotauluja ja palauttaa saman määrän samoin nimettyjä sarakkeita. UNION liittää kyselyiden tulokset suoraan toisiinsa. INTERSECT palauttaa tulosten leikkauksen, eli kaikki rivit, jotka esiintyvät molemmissa tuloksissa. EXCEPT palauttaa kaikki rivit, jotka esiintyvät vain toisessa kyselyistä. (PostgreSQL 2015, 7.4 Combining Queries.) Seuraava esimerkki yhdistää samassa taulussa esiintyvät korotetut ja maanalaiset kohteet.

```
SELECT *
FROM A
WHERE A.taso = 'korotettu'
UNION
SELECT *
FROM
WHERE A.taso = 'maanalainen'
```

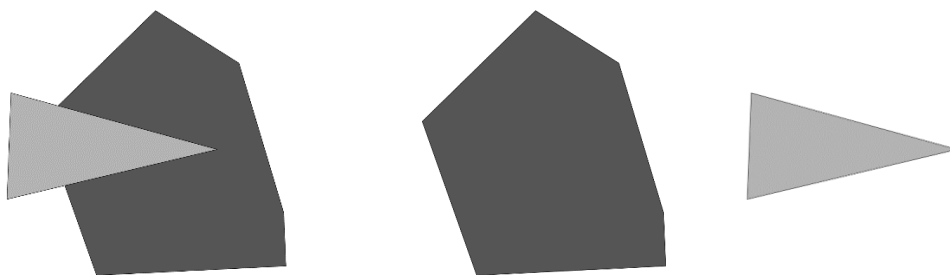
SELECT-rakenteen FROM-lauseessa valittuja tietotauluja voidaan yhdistää JOIN-avainsanalla. Liittämisehto määritellään ON-avainsanalla. CROSS JOIN yhdistää taulut A ja B siten, että jokaisella rivillä on aluksi taulun A sarakkeet, jonka jälkeen seuraa taulun B sarakkeet. Jos tauluilla on N ja M määrä rivejä, tuloksessa on rivejä yhteensä $N * M$. Huomioitavaa on, että FROM A CROSS JOIN B vastaa ilmaisua FROM A, B. Pelkkä JOIN tekee taulun, jossa on rivi jokaista B:n riviä kohden, jotka täyttävät liittämisehdon. LEFT JOIN:n tulostaulussa on rivi jokaista A:n riviä kohden. Aluksi suoritetaan JOIN, jonka jälkeen jokaista liittämisehtoa rikkonutta riviä kohden lisätään null-arvoja sisältävä rivi tulostauluun. RIGHT JOIN on käänteinen LEFT JOIN, eli tulostaulussa on rivi jokaista B:n riviä kohden. FULL JOIN on näiden yhdistelmä, eli tulostauluun tulee JOIN:n palautteen lisäksi null rivejä sekä LEFT- että RIGHT JOIN -funktioiden mukaisesti. (PostgreSQL 2015, 7.2.1 The FROM Clause.)

```
SELECT *
FROM A
JOIN B
ON A.id = B.id;
```

4.5 PostGIS-funktiot ja -operaattorit

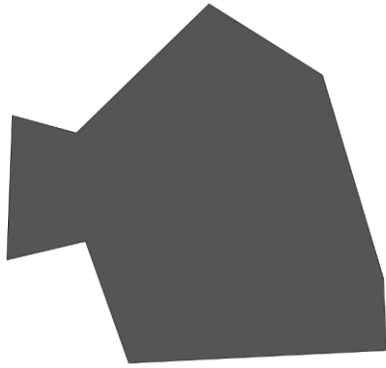
Tässä aliluvussa esitellään vain diplomityön käytännön osuuden kannalta olennaisia funktioita ja annetaan niiden käytöstä joitain esimerkkejä. Täysi lista PostGIS:n toiminnallisuudesta ja funktioiden täsmälliset käyttöohjeet löytyvät PostGIS:n teknisestä dokumentaatiosta (The PostGIS Development Group 2015).

PostGIS:n spatiaaliset funktiot perustuvat SFS ja SQL/MM-standardien määritelmiin, kuten ne on GEOS-moduulissa toteutettu. Spatiaaliset predikaatit on nimetty SQL/MM-standardin mukaisesti, ja niitä ovat ST_Equals, ST_Disjoint, ST_Touches, ST_Within, ST_Covers, ST_Intersects, ST_Contains, ST_CoveredBy, ST_Crosses ja ST_Overlaps. Nimetyt predikaatit voidaan myös korvata käyttämällä relaatioita vastaavia DE-9IM-lauseita ST_Relate-funktiossa. ST_Relate on nimettyjä relaatioita monipuolisempi, koska valmiit relaatiofunktiot ovat vain predikaatteja, eikä niiden avulla voida määritellä relaatien dimensioita, kuten DE-9IM sallisi. (The PostGIS Development Group 2015, 8.8 Spatial Relationships and Measurements.)



Kuva 9. Esimerkkipolygonit. Vasemmalla polygonit ovat limittäin, kuten niitä on prosessoitu seuraavissa esimerkeissä. Polygoni A on keskellä ja polygoni B oikealla.

ST_Union on niin kutsuttu sulautusoperaatio (engl. dissolve operation), joka ottaa syötteenä kohteiden geometrioita, ja sen tuloksena saadaan geometrisesti yhteen sulautettu kohde. Sulautumisessa kohteiden väliset rajat poistetaan ja kohteet liitetään yhteen, jolloin uudeksi kohteeksi muodostuu kuvan 9 alkuperäisten kohteiden peittämä ala kuvan 10 mukaisesti. Jos alaan kuuluu irrallisia kohteita, tuloksena saadaan samojen geometriatyyppeiden tapauksessa multikohde, tai erilaisia tyyppisiä sisältävän joukon tapauksessa geometriakokoelma. ST_Union-funktion kanssa käytetään usein ST_Dump funktiota, kun halutaan rikkoa multikohde yksittäisiksi kohteiksi. (The PostGIS Development Group 2015, 8.10 Geometry Processing.) ST_Union on hyödyllinen, mutta melko raskas operaatio, joka näkyy pitkinä suoritusaikoina suuren ja geometrisesti monimutkaisten aineistojen kanssa työskennellessä. Laatutestauksessa funktiota tarvitaan usein massiivistenkin polygoniaineistojen yhteydessä, jolloin suoritusajat saattavat venyä jopa useisiin tunteihin. Funktiota on pyritty kehittämään tehokkaampaan suuntaan, ja nykyversio suoriutuukin samasta tehtävästä ensimmäisiä toteutuksia merkittävästi nopeammin (Ramsey 2009). ST_Union esitetään visuaalisesti kuvassa n.n, jossa päällekkäin olevat kuvan n.n polygonit A ja B on sulautettu yhteen.



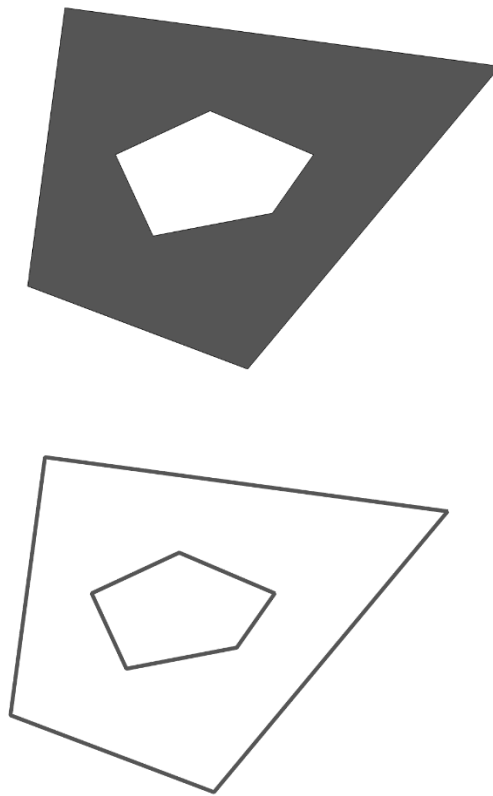
Kuva 10. Polygonit A ja B sulautettuna ST_Union-funktiolla.

ST_Intersection palauttaa sen kohdan, jossa kohteet leikkaavat. Polygonien tapauksessa kyseessä on alue ja viivojen tapauksessa piste. Pisteiden leikkaus on sama piste, koska leikkaavien pisteiden geometria on identtinen. (The PostGIS Development Group 2015.) ST_Intersection esitetään kuvassa 11.



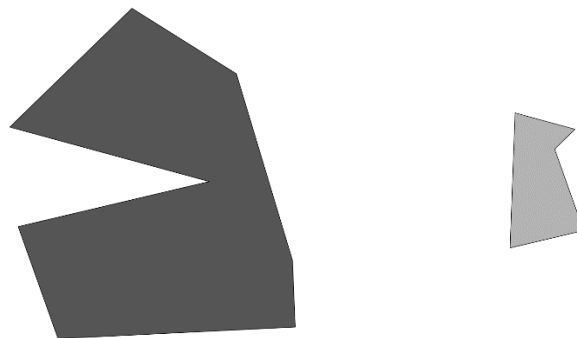
Kuva 11. Kohteiden A ja B leikkaus.

Geometrisiä kohteita voidaan myös tehdä analyysiä varten. Viivoista muodostetusta, suljetusta kehästä voidaan muodostaa polygoni ST_MakePolygon-konstruktorifunktiolla. Polygonin reunaviivat saadaan ST_ExteriorRing-funktiolla. Polygonin reikien reunaviivat saadaan ST_InteriorRingN-funktiolla. Funktiot toimivat vain yksittäisille polygoneille, joten multipolygonien yhteydessä joudutaan käyttämään ST_Dump-funktiota palauttamaan multipolygonin osat yksittäisiksi polygoneiksi (The PostGIS Development Group 2015, 8.4 Geometry Accessors). ST_MakeLine- ja ST_MakePoint-funktioiden avulla voidaan puolestaan tehdä viivoja tai pisteitä annettujen pisteiden ja koordinaattien perusteella. (The PostGIS Development Group 2015, 8.3 Geometry Constructors.) ST_BuildArea muodostaa polygonin tai multipolygonin annettujen kohteiden yhteisen reunaviivan perusteella, jonka tämä funktio pystyy itse muodostamaan kohteiden perusteella (The PostGIS Development Group 2015, 8.10 Geometry Processing). Polygonien reunaviivojen irroitus esitetään kuvassa 12.



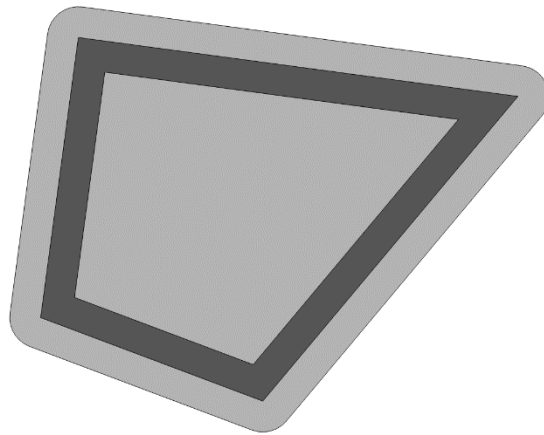
12: Sisä- ja ulkoreunojen irroitus. Ylempänä lähtötilanne ja alempana sisä- ja ulkoreunat, jotka voidaan täyttää polygoneiksi ST_MakePolygon-funktion avulla.

ST_Difference laskee geometrioiden erotuksen. Tuloksena saadaan kaikki sellaiset kohteiden osat, jotka eivät ole päällekkäin. Tyypillinen tulosjoukko koostuu osista alkupe-
räisten kohteiden reunoja sekä rei'istä toisen kohteen sisällä. Eli jos pienempi kohde ase-
tetaan suuremman kohteen päälle siten, että se on kokonaan suuremman kohteen sisällä,
saadaan näiden kohteiden erotuksena suurempi kohde, jossa on pienemmän kohteen ko-
koinen ja muotoinen reikä. Samoin jos viiva kulkee polygonin halki, saadaan erotuksena
polygonin puoliskot tai toisin päin viivan kärjet. ST_SymDifference tekee erotuksen sym-
metrisesti, jolloin operaation tuloksena saataisiin sekä polygonin puoliskot että viivan
kärjet. (The PostGIS Development Group 2015, 8.10 Geometry Processing.) Polygonien
erotus esitetään kuvassa 13.



*Kuva 13. Polygonien erotus. Vasemmalla polygoni B on vähennetty polygoni A:sta. Oi-
kealla polygoni A on vähennetty polygoni B:stä.*

Eräs yleisimmistä paikkatietoanalyysin menetelmistä on buffereiden käyttö. Bufferi on kohteen ympärille muodostettava alue, jonka avulla voidaan selvittää, onko jokin asia tietyn etäisyyden sisällä tai ainakin tämän etäisyyden ulkopuolella kohteesta. Positiivinen bufferi muodostetaan kohteen reunasta ulospäin määritellyn etäisyyden kattavaksi alueeksi. Negatiivinen bufferi muodostetaan kohteen reunasta sisäänpäin. Kuvassa 14 on kolme polygonia päällekkäin. Alin polygoni, jonka reuna nähdään kuvassa uloimpana, on alkuperäisen kohteen ympärille muodostettu positiivinen bufferi. Keskimmäinen polygoni, jonka reuna näkyy keskellä, on alkuperäinen polygoni. Päällimmäinen, pienin polygoni on taas alkuperäisen kohteen sisälle muodostettu negatiivinen bufferi. Bufferi voidaan muodostaa väliaikaiseksi kohteeksi ST_Buffer-funktiolla, tai samaan tapaan voidaan vaihtoehtoisesti tarkastella, ovatko kaksi kohdetta tietyn etäisyyden sisällä toisistaan ST_DWithin-funktiolla. (The PostGIS Development Group 2015, 8.8 Spatial Relationships and Measurements, 8.10 Geometry Processing.) ST_DWithin on yleensä suositeltavampi tapa selvittää, onko jokin asia tietyn etäisyyden sisällä kohteesta, sillä samaan tulokseen päästäkseen sen ei tarvitse muodostaa jokaisen tutkittavan kohteen ohessa uutta geometristä kohdetta (Ramsey 2007).



Kuva 14. Positiivinen ja negatiivinen bufferi samalle polygonille. Ulompi kehä kuuluu positiiviseen bufferiin ja tumma keskikohta alkuperäiseen polygoniin. Vaalea sisäosa on negatiivinen bufferi.

5 Laadunarvioinnin menetelmät ja työkalut projektissa

Tässä luvussa esitellään työn käytännön osuudessa käytetyt aineistot, menetelmät ja ohjelmistot sekä taustatietoa ELF-projektista.

5.1 Tavoitteet ja lähtötilanne

Diplomityön ensisijaisena tavoitteena oli selvittää, voidaanko PostGIS-tietokannassa olevan paikkatietoaineiston laatua arvioida avoimeen lähdekoodiin perustuvilla työkaluilla ELF-projektin laatuvaatimusten mukaisesti. Koska PostGIS vastaa toiminnallisuudeltaan käytännössä modernia paikkatieto-ohjelmaa (Ramsey 2010), lähdettiin toteutusta hakemaan lyhyen taustaselvityksen jälkeen suoraan PostGIS-funktioista. Menetelmän valtaisana etuna on, että laadun arviointia varten tietokannassa olevaa aineistoa ei tarvitse siirtää, eikä sille myöskään tarvitse tehdä minkäänlaisia konversioita. Testit ajetaan suoraan palvelimelle esitettävänä kyselyinä.

Kyselyiden tavoitteena oli saada niistä palautteeksi datasetti, jossa esiintyy kaikki yksittäistä laatusääntöä rikkovat kohteet. Koodin tyyliin tai suorituskykyyn ei kommentoimisen lisäksi ollut tarkoitus kiinnittää erityistä huomiota tässä vaiheessa. Käytettyjen PostGIS-funktioiden toimintaa pyrittiin kuitenkin tarkastelemaan jollain tasolla, mutta olennaisinta oli arvioida niiden soveltuvuutta ELF-projektin automaattisen laadunarvioinnin toteuttamiseen ja yleensä paikkatietojen laadun arviointiin.

5.2 Laadunarviointi ELF-projektissa

European Location Framework (ELF) -projektin tavoitteena on kehittää INSPIRE-yhteensopivien paikkatietojen välittämiseen soveltuva tekninen infrastruktuuri, joka palvelee koko Euroopan aluetta. ELF-aineistot ovat jokaisessa osallisessa maassa kansallisesti tuotettuja ja hallinnoituja, joten yhteiseurooppalaista käyttöä varten aineistot harmonisoidaan yhteensopiviksi maiden rajojen yli. Aineistojen laadunarviointi on osa tätä prosessia, sillä aineistot eivät saisi merkittävästi poiketa toisistaan myöskään laadullisilta ominaisuuksiltaan. ELF-aineistojen käyttäjinä toimivat niin kartta- ja ohjelmistotuotteiden loppukäyttäjät kuin aineistoja ja palveluita käyttävät ohjelmistokehittäjätkin. (Jakobsson 2012.)

Aineistontuottajat käyttävät kansallisten aineistojensa harmonisointiin projektia varten kehitettyjä ELF-geotyökaluja, joihin kuuluvat kolme laadunarviointityökalua, eli ArcGIS Data Reviewer, ppprepair ja 1 Validate. Määritelmien mukaan laadunvarmistustyökalun on määrä ottaa syötteenä kansallista ELF-dataa. Palautteena on tarkoitus saada datasetti, jossa näkyy aineiston virheet, laatusäännön tunnus, lyhyt kuvaus säännöstä sekä mahdollisesti myös muita virheeseen liittyviä kohteita. Lisäksi saadaan tilastotietoa muun muassa virheiden määrästä ja prosessointiajoista. Kansallinen data voi helposti poiketa muodoltaan muusta datasta, joten työkalut määritetään vakioparametreilla, jotka ovat vaivatonta muuttavissa käytettävän aineiston mukaan. (Werhahn et al. 2015.)

ISO 19139 -standardi tarjoaa menetelmiä metatietojen muodostukseen, ja tätä standardia käytetään ELF-metatiedoissa. Standardi kattaa kaikkien laatulementtien raportoinnin, mutta automaattiset laatu työkalut koskevat vain täydellisyyttä ja loogista eheyttä. Jokaisen laatulementin kohdalla ilmoitetaan ainakin tietotuotemäärittelyn tunnus, määrittelyiden version päiväys, selitys laadunarvioinnin tuloksesta sekä mahdollisesti linkki laatu-raporttiin. Yhtenevyys määrittelyihin ilmaistaan yhtenevyysasteella, joka on *true*, jos aineisto on yhtenevä, ja *false*, jos aineisto ei ole määrittelyiden mukainen. (Werhahn et al. 2015, Annex B.)

5.2.1 Aineistotuotteet ja -teemat

ELF-projektissa tietoaaineisto on jaettu omia teemojaan edustaviksi tietotuotteiksi INSPIREn määrittelyiden mukaisesti. Seuraavassa esitellään tietotuote aluksi englanninkielisellä nimellä, jonka jälkeen on suluissa merkitty tietotuotteen tunnus, ja tämän jälkeen englanninkielisen nimen suomennos. Tietotuotteiden lyhenteitä käytetään jatkossa ainakin ratkaisun esittelyssä. Valmiiden ELF-määrittelyiden puuttuessa, kaikki ELF-tietotuotteiden esittelyt perustuvat tässä työssä ensisijaisesti INSPIREn aineistodokumentaatioon (INSPIRE 2008), ja niiden suomenkieliset tulkinnat Euroopan Unionin direktiiviin 2007/2/EY (2007).

Addresses (AD), eli osoitteisto kertoo kohteen sijainnista osoitteen perusteella. Osoitteella voidaan tarkoittaa rakennuksen yksilöimistä tien nimen, talon numeron sekä postinumeron perusteella. Yksilöiminen voi kohdistua myös kiinteistöihin. (INSPIRE 2008, ELF Feature Catalog 2014.)

Administrative Units (AU), eli hallinnolliset yksiköt on tietoa hallintorajojen rajaamista alueista, joilla on jonkinlaisia paikallisia hallinto-oikeuksia. Kansalliset alueet, tai maat, jakautuvat hallinnollisiin yksiköihin, kuten esimerkiksi maakuntiin ja edelleen kuntiin tai kaupunkeihin. Hallinnolliset yksiköt kuvataan polygoneina ja niiden rajat viivoina. (INSPIRE 2008, ELF Feature Catalog 2014.)

Buildings (BU), eli rakennukset on tietoa ihmisen rakentamien kohteiden sijainneista ja geometriasta. BU on polygoniaineistoa, joten rakennukset ja niiden osat kuvataan polygoneina. Kohteiden kaksiulotteisen sijainnin lisäksi myös perustusten korkeuskoordinaatti sekä mahdollisesti rakennuksesta määritellyn kohdan, esimerkiksi katon harjan, korkeus maanpinnasta on myös ilmoitettu. (INSPIRE 2008, ELF Feature Catalog 2014.)

Cadastral Parcels (CP), eli kiinteistörekisteriyksiköt on tietoa kiinteistörekisteriin kirjatuista kiinteistöistä. Kiinteistöt kuvataan polygoneina ja niiden rajat viivoina. (INSPIRE 2008, ELF Feature Catalog 2014.)

Elevation (EL), eli korkeusmallit ovat digitaalisia korkeusmalleja maan-, jään- sekä merenpinnoille. Vesistöjen tapauksessa korkeusmalleihin sisältyy myös pohjanmuodot sekä rantaviivat. EL on rasteriaineistoa, jossa yksi ruutu vastaa maastossa olevan paikan korkeutta. (INSPIRE 2008, ELF Feature Catalog 2014.)

Geographical Names (GN), eli nimistö on paikannimihakemisto, joka sisältää muun muassa kaupunkien ja muiden asutusalueiden sekä luonnollisten tai historiallisten kohteiden paikannimiä. GN on pisteaineistoa, jossa nimetyn paikan sijaintia merkitään yksittäisellä pisteellä. (INSPIRE 2008, ELF Feature Catalog 2014.)

Hydrography (HY), eli vesistö sisältää tietoa erilaisista vesialueista, kuten meristä, järvistä ja joista. HY muodostuu kaikista kolmesta geometriatyypistä. Joet ovat merkitty viivoina, vesialueet polygoneina tai multipolygoneina. Pisteinä voidaan merkitä erilaisia erikoiskohteita, kuten patojen ja sulkujen paikkoja. (INSPIRE 2008, ELF Feature Catalog 2014.)

Transport Network (TN), eli liikenneverkosto on tietoa vesiväylistä, maantie- ja raitiotie verkoista sekä niihin liittyvästä infrastruktuurista. TN kuvaa väylät viivoina ja muut kohteet, kuten risteykset, yleisimmin pisteinä. Väylöillä tarkoitetaan teitä, raitioiteita ja vesiväyliä. Alueellisia kohteita, kuten lentokenttiä, merkitään polygoneina. (INSPIRE 2008, ELF Feature Catalog 2014.)

Protected Sites (PS), eli suojelualueet ovat alueita, jotka ovat suojeltuja kansallisten tai kansainvälisten säädösten perusteella. Suojelualueet muodostuvat polygoneista ja multipolygoneista. (INSPIRE 2008, ELF Feature Catalog 2014.)

Land Cover (LC), eli maanpeite kuvaa maanpinnan biologisia ja fysiologisia ominaisuuksia. Maanpeite on polygoniaineistoa, jossa polygonit ja multipolygonit edustavat maanpinnan erilaisia materiaalityyppejä, mutta eivät kerro alueiden käytöstä. Maanpeite voi siis olla esimerkiksi metsää, kalliota, ihmisen tekemiä kohteita tai vesialueita. (INSPIRE 2008, ELF Feature Catalog 2014.)

5.2.2 Laatusäännöt

ELF-projektin vielä kehittyvässä laatudokumentaatiossa laatusääntöjä oli kirjoitushetkellä yhteensä 105 kappaletta. ELF:n laatusäännöt v0.6 koskevat ainoastaan loogista eheyttä ja täydellisyyttä. (Werhahn et al. 2015, Annex A.) Tässä aliluvussa käydään läpi aineistoteemoja koskevia sääntökokonaisuuksia, jotka on listattu taulukossa 3. Laatusäännöt on yksilöity tunnuksen ja sen perään liitetyn numeron avulla, esimerkiksi GEO08 tai TRA14. Yksityiskohtaiset selostukset löytyvät kokonaisuudessaan ELF:n laatudokumentaatiosta (Werhahn et al. 2015, Annex A), sekä supistettuna liitteestä 1.

Yleiset laatusäännöt General (GEN ja DEF), ovat sääntöjä ja yleisiä määrittelyitä, jotka koskevat yleisesti kaikkia aineistoteemoja tai useampien teemojen ryhmiä, kuten kaikkia polygoniaineistoja. DEF määrittelee kaikkia teemoja koskevia käsitteitä, joihin viitataan aineistoteemoja koskevissa vaatimuksissa. Näitä ovat geometriatyyppien määrittelyt sekä rakojen ja säleiden sekä säleiden ohuussuhteen määrittelyt. GEN-säännöt asettavat kaikkia aineistoteemoja koskevia laatuvaatimuksia, jotka tulee kohdetyypistä riippuen tarkastaa jokaisen teeman kohdalla. Yhteensä kaikkia aineistoteemoja koskevia sääntöjä on 19 kappaletta. Suurimmalle osalle säännöistä pystyttiin laatimaan testit, mutta GEN16 jäi tässä vaiheessa tekemättä, sillä määrittely koskee 2,5-ulotteisen geometrian EPSG-koodia, eikä 2,5-ulotteista aineistoa ollut saatavilla. Lisäksi säännössä annettu EPSG-koodi on tavanomaisesta poiketen moniosainen. (Werhahn et al. 2015, Annex A.)

AdministrativeUnits-teeman säännöt (ADM) koskevat hallinnolliset yksiköt (AD) -teemaa, ja niitä on yhteensä 7 kappaletta. Säännöistä ensimmäinen koskee puuttuvaa tietoa ja loput kuusi topologista eheyttä. Täydellisyyttä testaavan säännön on tarkoitus varmistaa, että aineistossa yleensä on kohteita. Samankaltainen määrittely on jokaisen teeman yhteydessä. Koska kyseessä on polygoniaineistoa, topologista eheyttä koskevat säännöt liittyvät vierekkäisten polygonien suhteisiin, eli tässä tapauksessa hallintorajojen

eheyteen. Rajaviivojen pitää olla yhtenäisiä ja ilman turhia solmuja, eikä polygonien välissä saa olla aukkoja. Säännöissä havaittiin ristiriitaisuutta, sillä aukkoja koskevat säännöt kumosivat toisensa. ADM02 määrittelee, ettei saman nationalLevel-attribuutin arvon omaavien kohteiden välissä saa olla aukkoja. ADM09 kumoaa tämän säännön määrittelemällä, ettei minkään kohteiden välissä saa olla aukkoja. Laatudokumentatio ei ollut kirjoitushetkellä lopullinen versio, joten siinä saattoi olla virheitä. (Werhahn et al. 2015, Annex A.)

Buildings-teeman säännöt (BUI) koskevat rakennuksia kuvaavaa aineistoteemaa. Vaikka kyseessä on polygoniaineistoa, eivät teemalle yksilöidyt säännöt koske lainkaan kohteiden geometriaa. Säännöistä kolme koskevat arvojoukkoeheyttä sekä kaksi formaattieheyttä ja käsitteellistä eheyttä. (Werhahn et al. 2015, Annex A.)

Cadastral Parcels -teeman säännöt (CAD) koskevat kiinteistörekisteriyksiköitä. Sääntöjä on vain kaksi, ja molemmat koskevat topologista eheyttä. Polygoniaineistoille tyyppillisesti kohteiden välissä ei saa olla aukkoja, eikä niiden rajaviivoissa saa olla ylimääräisiä solmuja. CP sisältää valtavan määrän polygoneja, ja vaikka ne ovat verrattain yksinkertaisia, kestää koko aineiston geometriaa koskevissa testeissä kauan. (Werhahn et al. 2015, Annex A.)

GeographicalNames-teeman säännöt (GEO) koskevat nimistöä. Sääntöjä on yhteensä 9 kappaletta, joista 8 koskee käsitteellistä eheyttä ja yksi puuttuvaa tietoa. Kaikki teeman testit ovat taulukkokyselyitä, joissa tutkitaan ja vertaillaan sarakkeiden arvoja. Voidaan esimerkiksi testata, onko kohteen kieli ilmoitettu oikein. (Werhahn et al. 2015, Annex A.)

Hydrography-teeman säännöt (HYD) koskevat vesistöä. Kyseessä on teemojen määrittelyistä laajin 28 säännöllä, joista valtaosa, eli 19 kappaletta koskevat topologista eheyttä, 6 arvojoukkoeheyttä sekä 2 käsitteellistä eheyttä. Vain yksi sääntö koskee puuttuvaa tietoa. HY-aineistoteema sisältää kaikkia geometriatyyppejä, joten topologisen eheyden arvioinnissa geometriset testit koskevat usein useampaa geometriatyyppiä. Esimerkiksi voidaan testata, leikkaako rantaviiva vesialueen kanssa, tai onko joen solmukohdissa pisteitä. (Werhahn et al. 2015, Annex A.)

TransportNetwork-teeman säännöt (TRA) koskevat ilma-, vesi-, kaapeli-, raide- ja tie-liikenneverkostoja. Yksi 19:sta säännöstä koskee puuttuvaa tietoa ja loput 18 topologista eheyttä. Valtaosa topologista eheyttä koskevista säännöistä liittyy viivojen välisiin suhteisiin, sillä vaikka teema sisältää kaikkia geometriatyyppejä, on valtaosa aineistosta viivoja. TRA-sääntöjen puitteissa voidaan testata esimerkiksi ovatko kiitoradat lentokenttäalueen sisällä, ovatko satamat yhdistetty laivaväyliin tai ovatko raitiotiet yhtenäisiä. Testien suorituskyvyn kannalta erityisesti tieliikenneverkosto on hyvin massiivinen datasetti. (Werhahn et al. 2015, Annex A.)

ProtectedSites-teeman säännöt (PRO) koskevat suojelualueita. Sääntöjä on vain yksi, joka koskee puuttuvaa tietoa. Säännössä todetaan, että aineistossa pitää olla kohteita. (Werhahn et al. 2015, Annex A.)

LandCover-teeman säännöt (LAN) koskevat maanpeitettä. Sääntöjä on kaksi, joista ensimmäinen koskee ylimääräistä tietoa, ja toinen puuttuvaa tietoa. Kohteet eivät saa olla pinta-alaltaan liian pieniä, ja aineistossa pitää olla ainakin yksi kohde. Testien kannalta kyseessä on jälleen geometrialtaan valtaisan monimutkainen ja suuri datasetti, mutta raskaimmat testit koskevat yleistettyjä aineistoja. (Werhahn et al. 2015, Annex A.)

Cross-theme säännöt (CRO), eli ristikkäiset teemat edustavat sellaisia sääntöjä, joissa käsitellään useampaa teemaa samanaikaisesti. CRO-säännöissä ei testata, onko aineistossa kohteita, sillä se testataan jo teemojen omien sääntöjen yhteydessä. Voidaan esimerkiksi määritellä, että laivareittien pitää olla veden päällä tai asuinalueiden pois lentokentän päältä. (Werhahn et al. 2015, Annex A.)

Addresses- ja **Elevation-**teemojen laatusääntöjä ei ollut saatavilla lainkaan vielä dokumentin versiossa 0.6. Uutta versiota laatusäännöistä oltiin aikeissa ryhtyä suunnittelemaan, mutta uudistusten laajuudesta ei ollut tietoa kirjoitushetkellä. (Werhahn et al. 2015, Annex A.)

Taulukko 4. Laatusääntöjen jakautuminen teemoittain.

GEN	ADM	BUI	CAD	GEO	HYD	TRA	PRO	LAN	CRO
DEF07	ADM01	BUI01	CAD01	GEO05	HYD01	TRA01	PRO02	LAN01	CRO01
DEF08	ADM02	BUI02	CAD02	GEO06	HYD02	TRA02		LAN02	CRO02
DEF09	ADM03	BUI03		GEO07	HYD03	TRA03			CRO03
GEN02	ADM04	BUI04		GEO08	HYD04	TRA04			CRO04
GEN03	ADM06	BUI05		GEO09	HYD05	TRA05			CRO05
GEN04	ADM08	BUI06		GEO10	HYD06	TRA06			CRO06
GEN05	ADM09	BUI07		GEO11	HYD07	TRA07			CRO07
GEN06		BUI08		GEO12	HYD08	TRA08			CRO08
GEN07		BUI09		GEO13	HYD09	TRA09			CRO09
GEN08					HYD10	TRA10			CRO10
GEN09					HYD11	TRA11			CRO11
GEN11					HYD12	TRA12			CRO12
GEN12					HYD13	TRA13			
GEN13					HYD14	TRA14			
GEN14					HYD15	TRA15			
GEN15					HYD16	TRA16			
GEN16					HYD17				
GEN50					HYD18				
GEN80					HYD19				
GEN81					HYD20				
					HYD21				
					HYD22				
					HYD23				
					HYD26				
					HYD27				
					HYD28				
					HYD29				
					HYD31				

5.3 Kehitysympäristö ja työkalut

Työssä käytettiin PostgreSQL 9.4:ä, jolle konfiguroitiin PostGIS 2.1.7 -laajennus. Tietokantaa hallittiin PostgreSQL:n mukana asentuneella pgAdminIII-ohjelmalla sekä QGIS:n DB Managerilla. QGIS:stä käytettiin versiota 2.8.2 Wien. QGIS osoittautui helppokäyttöisyytensä ja kattavien ominaisuuksiensa myötä erinomaiseksi ohjelmaksi myös aineiston manipuloinnissa ja tarkastelussa.

5.3.1 Aineiston lataaminen tiedostoksi WFS-palvelusta

Paikkatietoaineistot ovat usein saatavilla erilaisista latauspalveluista. Web Feature Service (WFS) on usein käytetty palvelumuoto, joka voidaan konfiguroida lähettämään ELF-aineistoskeemojen mukaista GML-muotoista dataa.

WFS-palvelulle voidaan esittää kolmea erityyppistä kyselyä. GetCapabilities palauttaa metatietodokumentin palvelusta ja siinä käytettävistä WFS operaatioista ja parametreista. DescribeFeatureType palauttaa kuvauksen palvelun tukemista kohdetyypeistä. Tärkein kyselytyypeistä on kuitenkin GetFeature, joka palauttaa käyttäjän toiveiden mukaista aineistoa. GetFeature on kyselytyypeistä myös monimutkaisin, ja tyypillisesti kyselyt sisältävätkin runsaasti erilaisia parametreja, kuten rajoitelaatikkoja tai määritelmiä palautteen muodosta tai käytetystä WFS-palvelun versiosta. (OGC 09-025r1 2010.)

WFS-palvelua käytetään POST- tai GET-kyselyiden avulla. Kevyet ja pienet GET-kyselyt voidaan tehdä yksinkertaisesti internet-selaimen kautta. Koska selain yleensä kaatuu tai yhteys katkeaa suuria kyselyitä käsitellessä, suuremmat GET-kyselyt ja kaikki POST-kyselyt joudutaan tekemään muilla menetelmillä. GNU Wget on tähän tarkoitukseen hyvä työkalu. Wget on saatavilla useimmille käyttöjärjestelmille asennettavina paketteina, ja sitä voidaan käyttää joko graafisena ohjelmana tai komentorivin kautta. (GNU Wget 2015.) Wget lataa ”url”-kohdassa määritellyn wfs-kyselyn tuloksen file.xml-nimellä oletuskansioon seuraavalla komennolla.

```
wget "url" -O file.xml
```

Eri versioita käytettäessä WFS:stä, tulee huomioida se, että GET-kyselyiden logiikka on hieman erilainen eri versioissa. Esimerkiksi rajoitelaatikon koordinaatit määritellään jokaisessa versiossa eri tavalla.

```
http://visukartta02:8080/elf-wfs/services/elf-lod0bu?service=WFS&version=2.0.0&request=GetFeature&typenames=elf-lod0bu:Building&count=-1
```

Tässä WFS-palvelusta ladataan rakennusten pohjakuvia GetFeature-kyselyllä. Kyselyt ovat http-muotoisia. Aluksi on määritelty palvelun osoite ja nimi. Tämän jälkeen määritellään aineistoteema sekä palvelu, jonka jälkeen seuraa palvelun versio. Request-kohdassa määritellään kyselyn tyyppi, jonka jälkeen valitaan, mitä kohteita palvelusta ladataan. Lopuksi count-avainsanalla ilmaistaan haluttu kohteiden määrä, joka on tässä kyselyssä ilmoitettu -1:llä rajoittamattomaksi. Tällöin kysely palauttaa kaikki kohteet tietokannassa. (OGC 09-025r1 2010.)

5.3.2 QGIS aineiston tarkastelussa ja prosessoinnissa

Diplomityötä tehdessä kaikkien aineistoteemojen mukaisia, kansallisia aineistoja ei ollut saatavilla, joten testiaineistoa jouduttiin generoimaan. Itse tehty aineisto osoittautui myös äärimmäisen hyödylliseksi erityisesti monimutkaisempia laatutestejä kehittäessä, sillä kokonaiset datasetit ovat valtavia, eivätkä välttämättä sisällä havaittavia virheitä. Generoidun aineiston avulla tiedetään tarkasti, mitä voidaan odottaa saatavan testin tuloksena. Hyvin suppean ja rajallisen datasetin ansiosta suoritusaajat ovat lähes reaaliaikaisia, jolloin testin kehittäminen on oikean aineiston kanssa työskentelyyn verrattuna merkittävästi paljon nopeampaa välittömän palautteen ansiosta.

Quantum GIS (QGIS) -ohjelmiston avulla kehitykseen tarvittavia tapauksia voitiin generoida piirtämällä virheellisiä kohteita edit-tilassa. Myös aineiston osittelu onnistui erittäin helposti QGIS:n valintatyökalulla, sillä valitut kohteet voidaan tallentaa omaksi tasokseen Save as...-toiminnolla.

QGIS:n DB-managerin avulla voidaan hallita PostGIS-tietokantaa helposti. Tietokantaan voidaan luoda uusia skeemoja ja tauluja, ja esittää kyselyitä, joiden tulokset saadaan tasoksi tarkastelua varten QGIS:n. Kaikki auki olevat tasot voidaan DB-managerilla myös viedä tietokantaan. Jos aineistot ovat tiedostoina esimerkiksi GML- tai Shapefile-muodossa, ne voidaan lisätä tietokantaan yksinkertaisesti ensin avaamalla ne QGIS-tasoksi pudottamalla tiedostot QGIS-ikkunaan, ja viemällä ne edelleen DB-managerilla tietokantaan tauluiksi.

6 Laatutestit PostGIS-kyselyinä

Tässä luvussa käydään läpi esimerkkejä käytännön osuudessa toteutetuista paikkatietojen laatutesteistä täydellisyydelle ja loogiselle eheydelle. Jokaista laatulementtiä kohden on valittu joitain testien laatimisessa käytettyjä menetelmiä havainnollistavia tapauksia.

6.1 Täydellisyyden testaus

Täydellisyydessä on vain kaksi tapausta. Joko tietoja voi olla liikaa, tai niitä voi puuttua. ELF-laatuspesifikaatiossa täydellisyys koskee kohteiden minimikokoa, tai sillä tutkitaan, onko kohteita aineistossa lainkaan (Werhahn et al. 2015, Annex A).

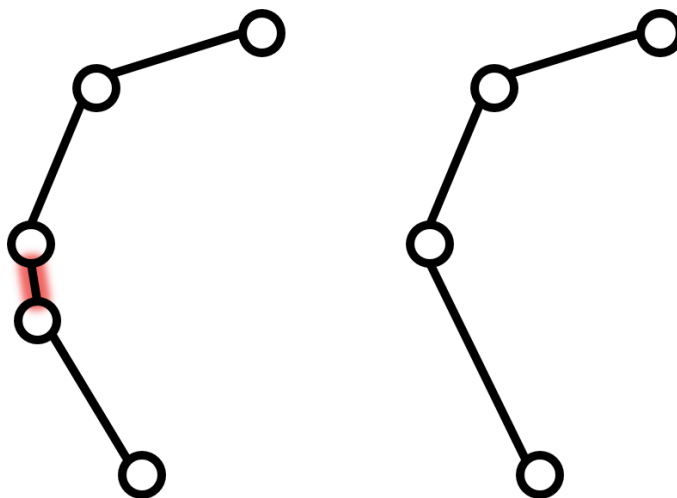
6.1.1 Kohteiden minimikoko

Suuremmille tarkkuustasoille siirryttäessä näytettävien kohteiden yksityiskohtaisuus ja määrä pienenevät. Tarkkuustasoaan varten alaltaan liian pienet polygonit on määritellyissä luokiteltu ylimääräiseksi tiedoksi. Virheelliset kohteet saadaan siis selvittämällä, mitkä kohteista ovat liian pieniä.

```
SELECT *  
FROM "LoD0_Buildings"."Building"  
WHERE ST_Area(geom) < 625;
```

GEN05-sääntöä vastaava kysely palauttaa kaikki rivit, joissa esiintyy polygoneja, joiden ala on pienempi kuin asetettu raja-arvo. Rakennusten minimikoko LoD2:ssa (Level of Detail, suom. tarkkuustaso) on 625m^2 , alueellisella tarkkuustasolla $0,04\text{km}^2$ ja globaalilla $0,4\text{km}^2$. Muille polygoneille LoD2:sta eteenpäin minimikoko on 2500m^2 . (Werhahn et al. 2015, Annex A.)

Samoin liian lyhyet viivojen segmentit luokitellaan ylimääräiseksi tiedoksi. Määrittely pätee viivoille ja myös polygonien reunaviivoille. Liian lyhyet viivojen segmentit tekevät kohteista tarpeettoman monimutkaisia suuremmilla tarkkuustasoilla. Kuvassa 16 esitetään liian lyhyen segmentin periaate visuaalisesti.



Kuva 16. Liian lyhyt segmentti viivassa. Vasemmanpuoleisessa viivassa liian lyhyt segmentti esitetään korostettuna. Oikeanpuoleinen viiva on määrittelyn mukainen.

GEN07-säännön mukaan viivoissa ja polygoneissa segmenttien pituus ei saa alittaa raja-arvoa, eli kahden rinnakkaisen solmun etäisyyden pitää olla suurempi tai yhtäsuuri kuin raja-arvo. LoD2:ssa raja-arvo on 1m, alueellisella tarkkuustasolla 5m sekä globaalilla tarkkuustasolla 50m. (Werhahn et al. 2015, Annex A.)

```

SELECT segments.*
FROM (SELECT id AS line_id, segment_id, geom, length,
      (CASE
        WHEN segments_sub.length < 1  --1m, 5m, 50m
        THEN true
        ELSE false
        END
      ) AS tooshort
FROM (
  SELECT a.id, segment_id, geom,
    ST_Length(
      ST_MakeLine(
        ST_PointN(geom, segment_id),
        ST_PointN(geom, segment_id + 1)
      )
    ) AS length
  FROM faultys.length AS a
  JOIN
    (SELECT id, generate_series(1, ST_NumPoints(geom)
      - 1) AS segment_id
    FROM faultys.length) AS b
    ON a.id = b.id
  ) AS segments_sub
) AS segments
WHERE tooshort IS TRUE;

```

Kysely etsii kaikki raja-arvoa lyhyemmät viivojen segmentit. Kyselyn avulla pyrittiin myös havainnollistamaan CASE-rakenteen toimintaa. Solmujen määrä saadaan ST_NumPoints-funktiolla, ja yksittäiseen solmuun viitataan ST_PointN-funktiolla (The PostGIS Development Group 2015). Viivan geometrian muodostamista varten viitataan kahteen peräkkäiseen solmuun. Virhekohteina palautetaan segmentit, joissa on geometrian lisäksi viittaus alkuperäisen viivan tunnuksen ja segmentin järjestyslukuun, segmentin pituus sekä kontrolliarvona totuusarvo siitä, onko kohde liian lyhyt vai ei.

6.1.2 Pakollisten kohteiden läsnäolo

Pakollisten kohteiden läsnäolo liittyy puuttuvaan tietoon. Virheelliset kohdetyypit saadaan selvittämällä, esiintyykö kohdetyyppejä edustavia kohteita lainkaan datasetissä. Määrittelyn tarkoituksena on saada raportti aineistoteemassa esiintyvistä kohteista (Werhahn et al. 2015, Annex A).

```

SELECT COALESCE(id)
FROM "AdministrativeUnits"."au-pl:AdministrativeUnit"
LIMIT 1;

```

ADM01-sääntöä vastaava kysely palauttaa ensimmäisen rivin tunnuksen, jonka id-sarake ei ole tyhjä. Kysely ei palauta mitään, jos datasetti on tyhjä. Kysely koskee vain yhtä saraketta, jotta se olisi mahdollisimman nopea.

6.2 Käsitteellisen eheyden testaus

Käsitteellisen eheyden testit ovat pääosin taulukkokyselyitä, joissa tutkitaan ja vertaillaan solujen arvoja. Käsitteellinen eheys tutkii, ovatko solujen arvot yhteneviä tietoaaineiston käsitteistön kanssa.

```
SELECT W.*
FROM "kadnl-hy"."Watercourse" W
WHERE W.origin = 'natural'
AND W.level = 'SuspendedOrElevated';
```

HYD28-määrittelyssä todetaan, että luonnolliset kohteet eivät saa olla korotettuja tai ripustettuja (Werhahn et al. 2015, Annex A). Korotettujen ja ripustettujen kohteiden oletetaan olevan esimerkiksi siltoja.

```
WITH SettlementAreaPoint AS (
    SELECT S.*
    FROM "GeographicalNames"."NamedPlace" S,
         "LandCover"."LandCoverUnit" LAN
    WHERE S.type = 'populated Place'
    AND L."LandcoverObservation" = 'artificial Construction'
    AND ST_Area(L.geom) >= 400 --0.4km2 @regional/global
    AND ST_Within(S.geom, L.geom)
)
SELECT L.*
FROM SettlementAreaPoint L
LEFT JOIN (
    SELECT P.*
    FROM SettlementAreaPoint P,
    WHERE P.population >= 5000 --5000 @regional, 50 000 @global
) R
ON R.id = L.id
WHERE R.id IS NULL;
```

CRO12-määrittelyn mukaan asuinalueiksi määritellyillä LandCoverUnit-kohteilla pitää olla minimiala tai minimimäärä asukkaita. Alueiden koon pitää olla yli 0,4km². Alueiden väkiluku tarkastetaan niiden sisällä olevista paikannimipisteistä, jotka on määritelty asuiksi paikoiksi ja niiden populaatio ylittää asetetun raja-arvon. Kysely palauttaa sellaiset asutuspisteet, joiden väkiluku on vähemmän kuin 5000. (Werhahn et al. 2015, Annex A.)

```
SELECT *
FROM "LoD0_Buildings"."Building"
WHERE ST_IsEmpty(geom)
OR ST_IsEmpty(ST_Envelope(geom));
```

GEN04-määrittelyssä todetaan, että kohteen geometria tai sen minimirajoitelaatikko ei saa olla tyhjä (Werhahn et al. 2015, Annex A). Kohteen geometria pitää siis olla olemassa. Kysely palauttaa kohteet, joiden geometria on tyhjä, eli sitä ei ole lainkaan.

6.3 Arvojoukkoeheyden testaus

Arvojoukkoeheys tutkii, kuuluvatko solujen arvot määrittelyissä sallittuihin arvoihin.

```
SELECT W.*
FROM "kadnl-hy"."Watercourse" W
WHERE W.width.lower >= W.width.upper;
```

HYD26 määrittelee, että jokien leveyden alarajan täytyy olla pienempi kuin sen yläraja (Werhahn et al. 2015, Annex A.) Kysely palauttaa kaikki joet, joiden alaraja on suurempi tai yhtäsuuri kuin sen yläraja.

```
SELECT *
FROM "LoD0_Buildings"."Building"
WHERE "horizontalGeometryEstimatedAccuracy" > 1000;
```

BUI09 määrittelee, että rakennusten arvioitujen korkeuksien tarkkuuden pitää olla pienempi tai yhtä suuri kuin raja-arvo (Werhahn et al. 2015, Annex A.) Kysely palauttaa kaikki raja-arvoa epätarkemmiksi arvioituja korkeuksia sisältävät kohteet.

6.4 Formaattieheyden testaus

Formaattieheydellä varmistetaan, että aineisto on formaatin mukaista, eli noudattaa sille asetettuja muodollisia vaatimuksia. ELF-laatuvaatimuksissa säännöt koskevat esimerkiksi oletettua koordinaattijärjestelmää tai standardien noudattamista.

```
SELECT *
FROM "GeographicalNames"."NamedPlace"
WHERE NOT ST_SRID(geom) = 4258;
```

GEN02 määrittelee, että ETRS89:n piiriin kuuluvalla alueella koordinaattijärjestelmänä tulee käyttää ETRS89-GRS80-järjestelmää, jonka EPSG-koodi on 4258. Vastaavasti tämän ulkopuolella tulee käyttää ITRS:ää, jonka viimeisimmän realisaation, ITRF2008:n EPSG-koodi on 5332. (Werhahn et al. 2015, Annex A.) Kysely palauttaa kaikki sellaiset nimistön kohteet, jotka eivät ole oikeassa koordinaattijärjestelmässä.

```
SELECT *
FROM "GeographicalNames"."NamedPlace"
WHERE NOT ST_IsSimple(geom);
```

GEN03 määrittelee, että kaikkien paikkatietokohteiden tulee noudattaa OGC 06-103r4 Simple Features -standardia (Werhahn et al. 2015, Annex A.) Kysely palauttaa kaikki sellaiset kohteet, jotka eivät noudata tätä sääntöä.

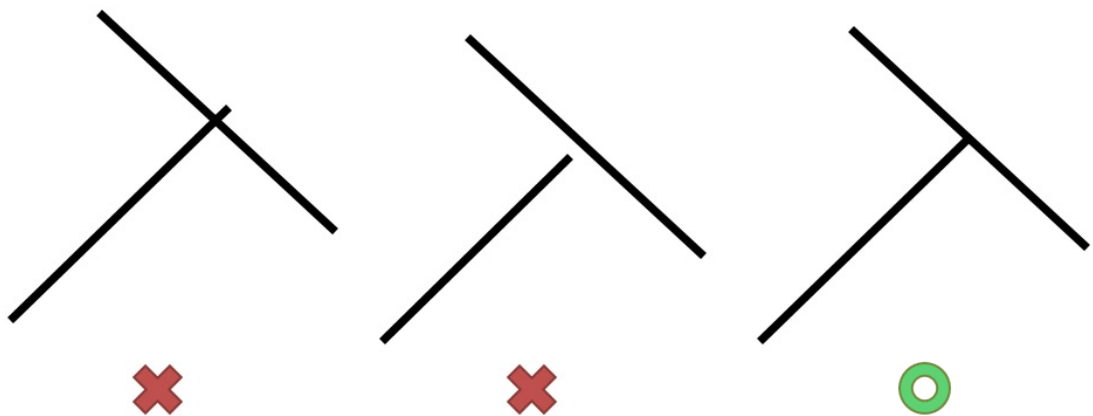
6.5 Topologisen eheyden testaus

Topologisen eheyden testit vaativat ylivoimaisesti eniten prosessointiaikaa, sillä ne koskevat aineiston geometrian validointia. Geometrian validointi vaatii paljon prosessointia, sillä kohteet saattavat olla geometrisiltä ominaisuuksiltaan hyvinkin monimutkaisia. Prosessointi tapahtuu tyypillisesti laskennallisen geometrian algoritmien avulla, jotka muistuttavat paljon tietokonegrafiikasta ja konenäöstä tuttuja menetelmiä. Algoritmit kuuluvat spatiaalisten funktioiden ja operaattoreiden sisäisiin toteutuksiin, joten niitä ei käsitellä tässä työssä.

Tapauksia on runsaasti, sillä valtaosa määrittelyistä koskee juuri topologista eheyttä. Kaikissa testeistä olennaisessa osassa ovat spatiaaliset relaatiot, sillä ELF-määrittelyissä topologinen eheys liittyy juuri geometrinen kohteiden suhteisiin toisiinsa nähden sekä niiden sisäisen geometrian eheyteen (Werhahn et al. 2015, Annex A).

6.5.1 Viivojen yhdistyvyys

Viiva-aineistoissa tyypillinen virhe on, etteivät viivat yhdisty oikein. Viivojen päät saattavat mennä virheellisesti yhdistymiskohdasta joko yli, tai ne saattavat jäädä liian lyhyeksi. Usein kyse on digitoinnissa tapahtuneesta virheestä. ELF:n määrittelyissä yhdistyvyys määritellään seuraavasti. Viivojen päätepisteiden välinen etäisyys tulee olla raja-arvoa suurempi, jos viivat L ja M ovat samaa tyyppiä, ja kumpikaan L:n päätepisteistä ei leikkaa M:n geometriaa missään kohdassa (Werhahn et al. 2015, Annex A). Viivojen yhdistyvyyden säännöt esitetään kuvassa 17.



Kuva 17. Viivojen yhdistyvyys. Tapaukset ovat vasemmalta oikealle overshoot, undershoot ja ei virhettä.

```
SELECT L.*
FROM faultys.lines L, faultys.lines M
WHERE L.id != M.id
AND ST_Disjoint(ST_EndPoint(L.geom), M.geom)
AND ST_Disjoint(ST_StartPoint(L.geom), M.geom)
AND (ST_DWithin(ST_EndPoint(L.geom), M.geom, 5)
OR ST_DWithin(ST_StartPoint(L.geom), M.geom, 5)
);
```

Määrittelyä GEN09 vastaan tehty kysely palauttaa kaikki sellaiset viivat, joiden päätepisteet eivät yhdisty oikein. Kyselyssä käytettiin ST_EndPoint- ja ST_StartPoint-funktioita viivojen päiden löytämiseen. Kysely on tarkoitettu vähintään LoD2-aineistolle, joten testattaessa LoD1-aineistolla tuloksena saatiin runsaasti virheitä.

6.5.2 Polygonien väliset aukot ja säleet

Aukot ja säleet ovat todennäköisesti tyypillisimpiä ongelmia polygoniaineistoissa. Aukolla tarkoitetaan kahden tai useamman polygonin väliin jäävää tyhjää tilaa. Yleensä polygoniaineistoissa ei saa olla aukkoja kohteiden välissä. Aukkoa kutsutaan säleeksi, kun se on havaittu tarpeeksi ohueksi. Aukon ohuus T määritellään ohuussuhteena, joka ELF-projektin vaatimusmäärittelyissä lasketaan kaavan 15 mukaisesti. Kaavassa A tarkoittaa

kohteen alaa ja p kohteen ympäröimittaa. Aukko on säle, jos sen ohuussuhde T on pienempi tai yhtäsuuri kuin 0,3. (Werhahn et al. 2015, Annex A.)

$$T = \frac{4\pi \cdot A}{p^2} \quad (15)$$



Kuva 17. LandCoverUnit-polygoniaineistoa Hollannin rannikolta. Koko datasettiin kuuluu noin 2,5 miljoonaa polygonia.

Kuten kuvasta 17 nähdään, polygoniaineisto on geometrisiltä ominaisuuksiltaan varsin monimutkaista. Silmämääräisesti tarkasteltuna voidaan todeta, että aineistossa on aukkoja. LC-aineistossa saa olla aukkoja, mutta säleitä ei saa olla lainkaan (Werhahn et al. 2015, Annex A). Pieniä säleitä ei voida löytää millään järkevällä menettelyllä silmämääräisesti, joten tämä tapaus on erittäin hyvä esimerkki siitä, että tyypillisesti ainoa vaihtoehto loogisen eheyden arviointiin on automaattinen laaduntarkastus siihen soveltuvalla ohjelmistolla. PostGIS:llä voidaan löytää aukkoja ja säleitä polygoniaineistossa esimerkiksi seuraavilla menettelyillä.

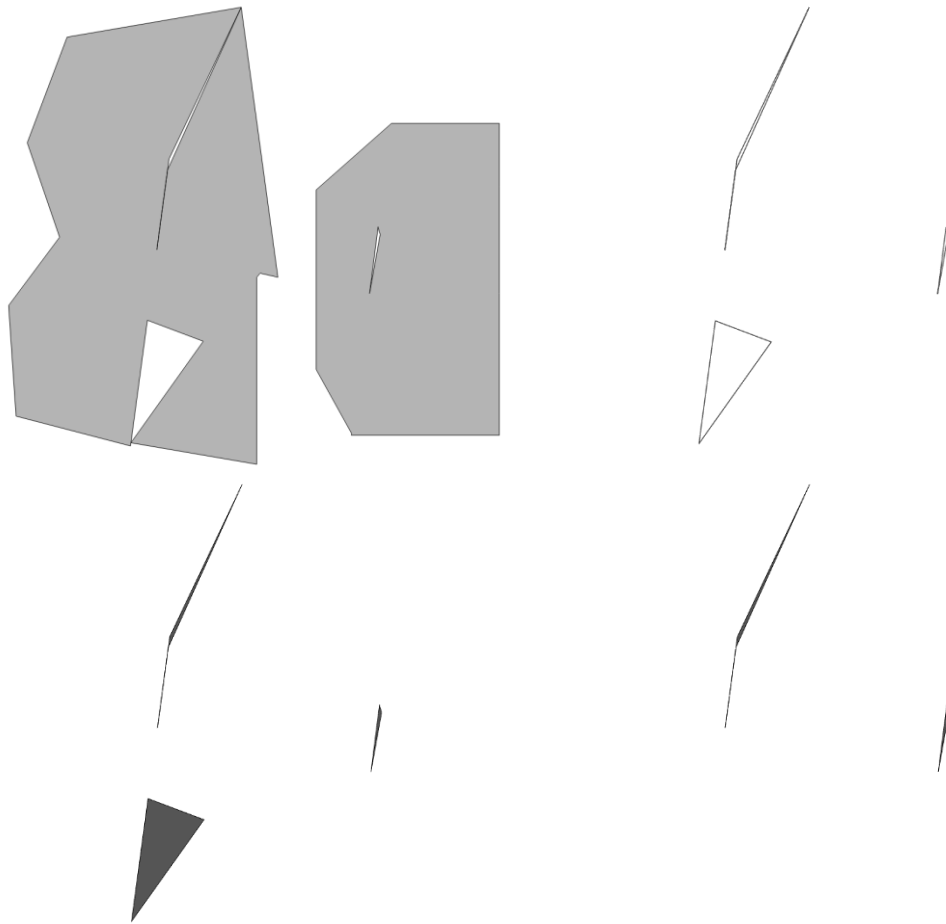
Aineiston kaikista polygoneista tehdään aluksi yhtenäinen polygoni $P1$ ST_Union -funktiolla. ST_Union suorittaa niin sanotun sulatusoperaation polygoneille (engl. dissolve operation), jossa polygonit yhdistetään siten, että niiden väliset rajat poistetaan. Polygoneista muodostuu siis yksittäinen $ST_Polygon$, jos kaikki polygonit koskettavat toisiaan. Jos kaikki polygonit eivät kosketa toisiaan, operaation tuloksena saadaan $ST_MultiPolygon$. Usein on niin, että menetelmä palauttaa multipolygoneja. Jotta seuraavat kohdat onnistuvat, joudutaan käyttämään ST_Dump -funktiota, joka palauttaa multipolygonin

osat yksittäisinä polygoneina. ST_Union funktion sijaan operaatioon voidaan käyttää myös ST_Collect- ja ST_UnaryUnion-funktioita, joista ensimmäinen kerää kaikki polygonit yhdeksi setiksi, ja seuraava sulattaa ne yhteen. Tällä vaihtoehdoisella menettelyllä voidaan joissain tapauksissa päästä parempaan suorituskyykyyn.

Menetelmässä A aukot löydetään polygonin geometrian määritelmän perusteella, sillä aukot kuuluvat polygonin geometriaan. Aukot voidaan laskea ST_NumInteriorRings-funktiolla, mikä mahdollistaa yksittäisten aukkojen ulkoreunojen muodostamisen N:nnen järjestysnumeron aukkoon viittaavalla ST_InteriorRingN-funktiolla. Aukkojen reunoista voidaan edelleen muodostaa polygoneja ST_MakePolygon-funktiolla. (Kania 2013.)

```
WITH uni AS ( --koko alueen kattava polygoni
    SELECT ROW_NUMBER() OVER() AS id,
           dump.geom geom
    FROM (
        SELECT (ST_Dump(ST_Union(geom))).geom geom
        FROM "LandCover"."amster" --change this for your data
    ) dump
),
num AS ( --aukkojen numerointi
    SELECT id polynum, ST_NumInteriorRings(geom) holenum
    FROM uni
),
n AS ( --aukkojen numerot liitetty vastaaviin polygoneihin
    SELECT polynum polynum, generate_series(1, holenum) holenum
    FROM num
),
gaps_pre AS ( --aukot
    SELECT ST_MakePolygon(ST_InteriorRingN(uni.geom, holenum))
           AS geom
    FROM n, uni
    WHERE id = polynum
),
gaps AS ( --aukot ja ohuussuhde
    SELECT *,
           (4*Pi()*ST_Area(geom))/(ST_Perimeter(geom))^2 AS thinness
    FROM gaps_pre
)
--valitaan virheet
SELECT ROW_NUMBER() OVER() AS id, *
FROM (
    SELECT *
    FROM gaps
    WHERE gaps.thinness <= 0.3 --thinness ratio is always 0.3
    AND ST_Area(geom) < 50 --LoD2: 50m2, Regional: 200m2,
Global: 500m2
) slivers;
```

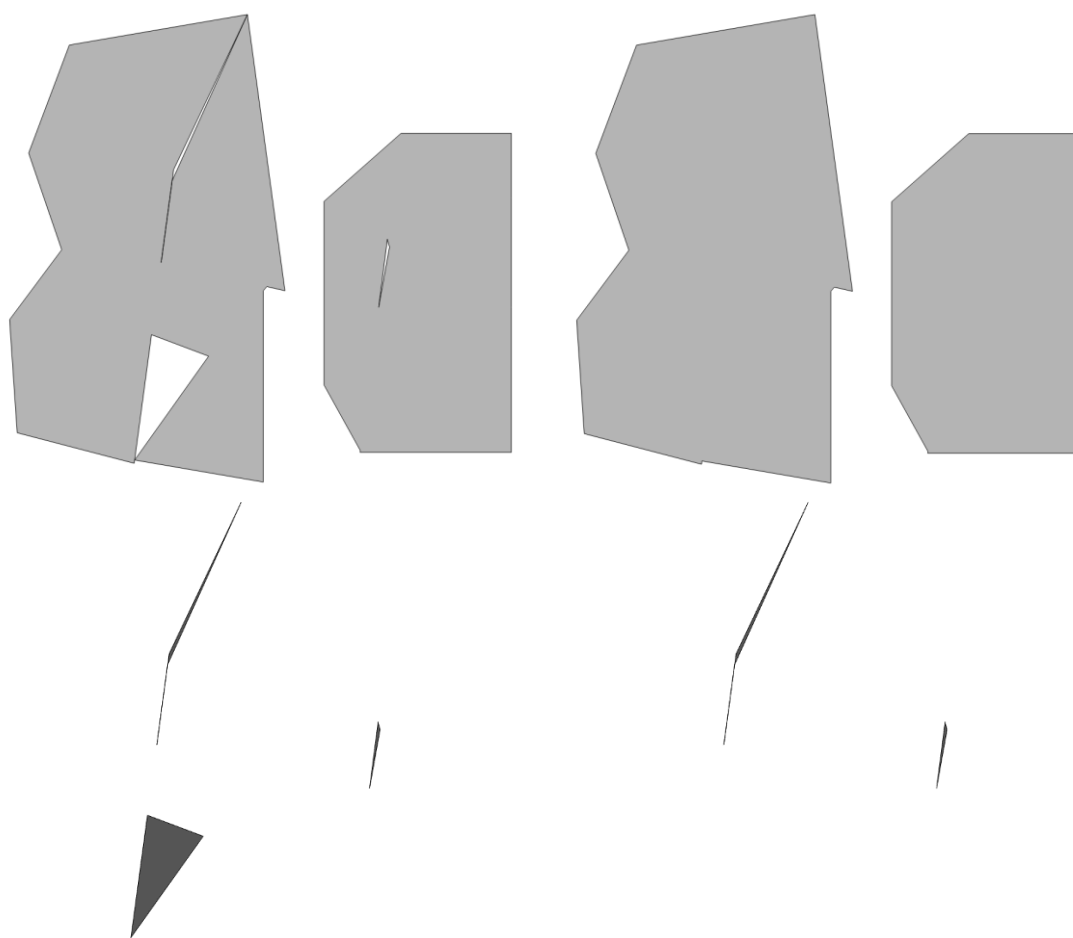
DEF09-määrittelyä ja GEN13-sääntöä vastaava kysely palauttaa kaikki sellaiset aukot, jotka todetaan säleiksi ohuussuhteen perusteella. Sälepolygonien riveille on merkitty säleelle annettu id, säleen geometria sekä ohuussuhde. Muita tietoja ei säleille ole tässä vaiheessa annettu, sillä säleet ovat täysin uusia polygoneja, jotka on muodostettu aukkojen reunaviivojen perusteella. Kuvassa 18 esitetään ratkaisu kuvina.



Kuva 18. Menetelmä A aukkojen ja säleiden löytämiseksi. Vasemmalla ylhäällä on sulautettu polygoni ja oikealla aukkojen reunaviivat. Alhaalla vasemmalla aukkojen reunaviivat on täytetty polygoneiksi ja oikealla on niistä poimitut säleet.

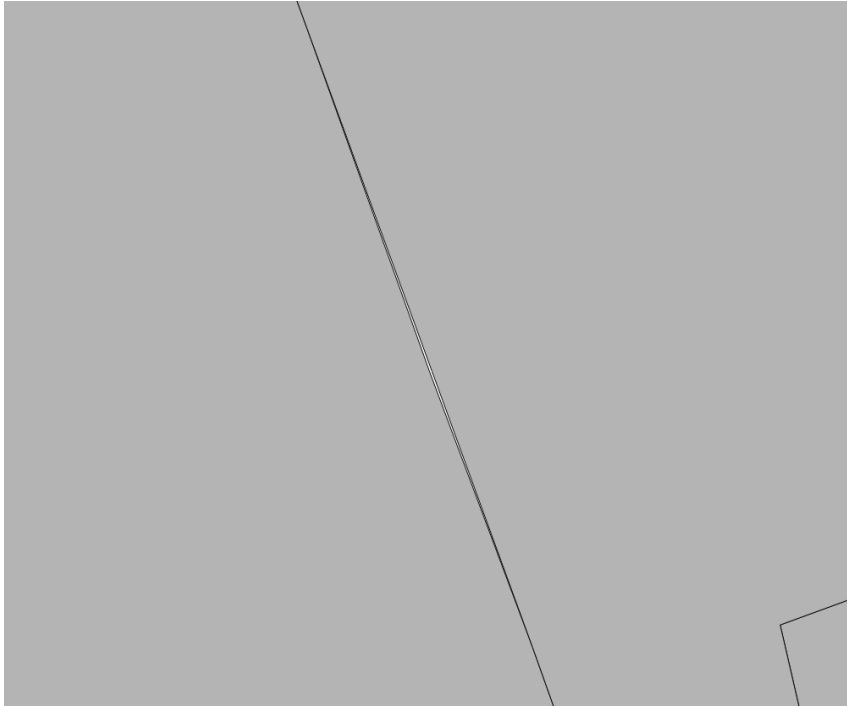
Vaihtoehtoisella menetelmällä B aukot löydetään erotuksen kautta. Aluksi otetaan polygonista P1 ulkoreunat ST_ExteriorRing-funktiolla, ja muodostetaan siitä polygoni P2 käyttäen jälleen ST_MakePolygon-funktiota. Aukot saadaan, kun lasketaan P1:n ja P2:n erotus ST_Difference-funktiolla. (Barça 2015.) Kuvassa 19 esitetään menetelmän eteneminen visuaalisesti. Erotuksen kautta aukot saadaan siis seuraavasti.

```
(ST_Dump(
  ST_Difference(
    ST_MakePolygon(
      ST_ExteriorRing(
        (ST_Dump(uni.geom)).geom
      )
    ), --ulkokuoren määrittämä polygoni
    uni.geom --alkuperäinen polygoni
  )
)).geom;
```

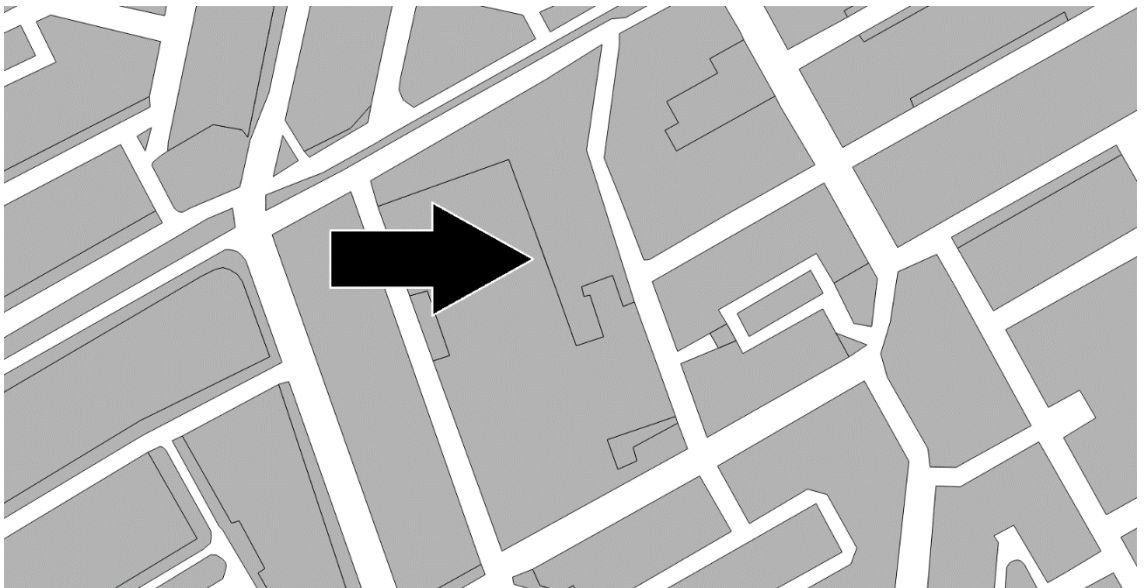


Kuva 19. Menetelmä B aukkojen ja säleiden löytämiseen. Vasemmalla ylhäällä on sulautettu polygoni ja oikealla sen ulkoreunasta muodostettu polygoni, jossa ei ole reikiä. Vasemmalla alhaalla on löydetty aukot ja oikealla säleiksi havaitut aukot.

Kuvassa 20 nähdään eräs löydettyistä säleistä. Löydetty säle on hyvin tyypillinen tapaus, joka sijaitsee kahden polygonin välissä. Säleen siro olemus viitanee digitoituvaiheen virheeseen.



Kuva 20. Hyvin pitkänomainen säle Hollannin LandCoverUnit-datasetissä.



Kuva 21. Ongelmakohdan sijainti hieman kauempaa tarkasteltuna. Kuvan 6.5.2b säleen paikka on osoitettu nuolella.

LandCover-aineistolla menetelmän A havaittiin olevan merkittävästi vahvempi kuin menetelmä B. Jos Amsterdamin kulman säleet löydettiin noin 15 minuutissa menetelmällä B, menetelmä A selvisi samasta operaatiosta peräti 10 sekunnissa. Koko Amsterdamin laajuisen otoksen käsittelyyn menetelmällä B kului useita tunteja, missä menetelmä A ratkaisi pulman noin 15 minuutissa. Näillä tiedoilla menetelmää B ei kokeiltu lainkaan koko Hollannin aineiston kanssa. Menetelmä A selvitti koko aineiston säleet noin 17,5 tunnissa. Kaikki suoritusajat on kerätty taulukkoon 4. Kuvat 20 ja 21 havainnollistavat säleiden olemusta osana datasettiä. Säleitä ei ole mahdollista löytää silmämääräisesti millään järkevällä menettelyllä, koska niiden havaitseminen edellyttää valtavan suuren aineistomäärän tarkastelua aivan lähietäisyydeltä.

Taulukko 5. menetelmien A ja B suoritusaikojen vertailu Amsterdamin osalle, Amsterdammille sekä koko Hollannille

Datasetti	A: ST_InteriorRingN	B: ST_Difference
Amsterdam osa	10 s	15 min
Amsterdam	15 min	tunteja
Hollanti	17,5 h	-

7 Johtopäätökset

Tässä työssä tutkittiin, voidaanko ELF-projektin laatuvaatimusmäärittelyihin perustuva paikkatietojen automaattinen laadunarviointi toteuttaa avoimen lähdekoodin ohjelmistoilla. Tutkimuskysymys pyrittiin todistamaan toteuttamalla kaikki laatusääntöjen mukaiset laatutestit PostGIS-tietokantaohjelmistolla. Tutkimuskysymyksen todistamisen ehdoksi asetettiin, että kaikille määrittelyille on vähintään selkeästi tunnistettavissa niiden toteuttamiseen tarvittavat PostGIS-funktiot ja kyselyn rakenne.

Työn tulokset esitetään liitteessä 1. Taulukon alussa on alkuperäisistä laatuvaatimusmäärittelyistä laatutestien kannalta olennaiset sarakkeet, jonka jälkeen taulukkoon on merkitty kaikki testeissä käytetyt PostGIS-funktiot ja kommentti testin suorituksesta. Kommenteissa *query done* tarkoittaa, että testiä varten on onnistuneesti laadittu virheelliset paikkatietokohteet palauttava kysely. Mikäli testin kehitystä varten ei ole ollut oikeaa aineistoa saatavilla, on kommenttiin merkitty *no data*. Ongelmallisten testien yhteydessä käytetään ilmaisua *issue*, sekä selitystä ongelman luonteesta. Myös muita sanallisia täsmennyksiä saatetaan antaa yksittäisten testien yhteydessä.

Tulosten perusteella voidaan todeta, että PostGIS-tietokantaohjelmisto kattaa kaikki paikkatietojen automaattisessa laadunarvioinnissa tarvittavat toiminnallisuudet. PostGIS osoittautui paikkatietoanalyysiominaisuuksiltaan jopa niin monipuoliseksi, että sen voidaan sanoa vastaavan mitä tahansa modernia paikkatieto-ohjelmaa. Kaikki ELF-laatuvaatimusmäärittelyiden mukaiset laatutestit voidaan toteuttaa PostGIS-kyselyinä. Kaikkien testien toimintaa ei pystytty kokeilemaan oikean datan kanssa, koska ELF-aineistoskeemojen mukaista dataa oli työn edetessä saatavilla ainoastaan rajallinen määrä. Lisäksi suuri osa puuttuvasta datasta selittyy yleistettyjen aineistojen puuttumisella, mutta usein testin toimintaa voitiin kokeilla tarkemman aineiston avulla. Noin puolet testeistä kehitettiin ELF-datan perusteella, jonka puuttuessa ratkaisuun päästiin jonkin muun aineiston, konstruoidun datan tai päättelyn avulla.

Täysin validiin ratkaisuun päästäisiin ainoastaan kehittämällä kyselyitä määrittelyjen mukaisten aineistojen kanssa. Tästä huolimatta tuloksia voidaan pitää pätevänä, sillä monien määrittelyiden mahdolliset toteutukset ovat keskenään hyvin samankaltaisia, ja siten myös toisiaan vastaavien kyselyiden käyttäytyminen on hyvin ennakoitavissa. Lisäksi testien toteutus ja tarvittavat funktiot voidaan päätellä muiden testien tai spatiaalisten reaalioiden avulla melko luotettavasti.

Testien suoritusajat kasvoivat joissain tapauksissa pitkiksi, mutta niistä saatiin onnistuneesti palaute myös suurten ja geometrisiltä ominaisuuksiltaan monimutkaisten aineistojen yhteydessä. Tulee toki huomioda, että yleistetyllä aineistolla topologisen eheyden testien suoritus nopeutuu, koska silloin aineiston geometriset ominaisuudet ovat yksinkertaisempia ja kohteita on vähemmän. Useat raskaiksi osoittautuneet testit ovat tarkoitettu yleistetyimmille aineistoille. Monien testien kehittämistä varten generoitiin tarkoituksenmukaisia datasettejä keinoitekoinen virheineen, mikä myös nopeutti testien kehittämistä merkittävästi välittömän palautteen ansiosta. Toteutuksen luotettavuus on aivan eri luokkaa, jos voidaan käytännössä osoittaa, että tuloksena tosiaan saadaan kaikki laatuvaatimusta rikkovat kohteet.

Suurin epävarmuus testeissä on niiden suorituskky. Testien suorituskkyä ja palautetta tulisikin verrata muihin laadunarviointityökaluihin, jotta PostGIS:n kyvyt saataisiin paremmin selville. Kyselyiden suorituskkyä arvioitiin ja parannettiin vain sellaisten kyselyiden yhteydessä, jotka osoittautuivat erityisen ongelmallisiksi. Suorituskkyä voidaan tutkia kyselyn eteen asetettavalla, kyselyn toimintaa ennen suoritusta arvioivalla EXPLAIN-avainsanalla, sekä kyselyn toimintaa suorituksen jälkeen analysoivalla ANALYZE-avainsanalla (Ramsey 2007). Erityisesti tietotauluja halutaan käydä läpi indeksihauulla (engl. index scan) ja koko taulukkoa läpikäyviä peräkkäishakuja (engl. sequential scan) halutaan välttää.

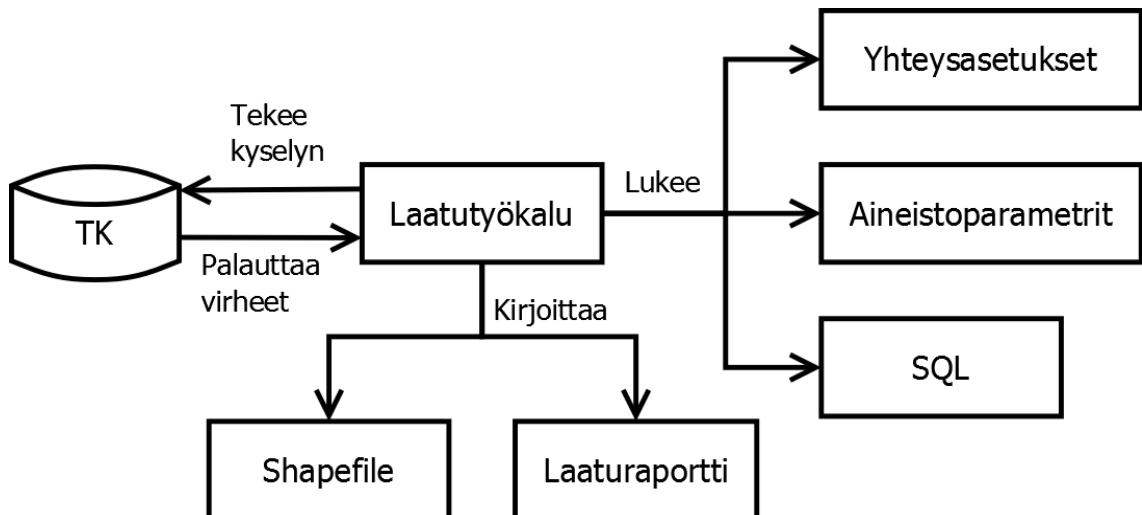
Testien optimoimiseksi kaikkea turhaa tai ylimääräistä laskentaa tulee välttää. Jos halutaan arvioida koko aineistoteeman laatua, voitaisiin suoritusajkoja vähentää siten, että jos useammassa testissä esimerkiksi lasketaan polygonien unioni koko aineiston geometrialle, se laskettaisiin vain kerran, eikä jokaiselle sitä käyttävälle testille erikseen. Erityisen raskaiksi operaatioiksi todettiin ST_Union ja ST_Intersection, mutta tutkimuksen perusteella niiden optimoimiseksi saattaa olla joitain toimivia keinoja, kuten ST_Unionin suorituksen pilkkominen pienempiin osiin.

Kyselyitä kehittäessä havaittiin, että kyselyä ajettaessa postgresql.exe käyttää tasan 25 % suorittimen resursseista. Tutkimustyön perusteella paljastui, että PostgreSQL voi käyttää yksittäiseen tietokantayhteyteen vain yhtä prosessoriydintä. PostgreSQL on suunniteltu useiden käyttäjien samanaikaista käyttöä ajatellen, joten yhden yhteyden moniydinominaisuuksia ei ole katsottu kovin tärkeiksi. Koko tietokantaa koskevia ja laskentaintensivisiä kyselyitä ajettaessa tämä muodostuu ongelmaksi. (PostgreSQL wiki 2015.) Asiaan on todennäköisesti kuitenkin tulossa korjaus jo vuoden 2016 syksyn aikana PostgreSQL 9.6:n myötä, sillä moniydintukea ollaan parhaillaan kehittämässä (Odoo.com forum 2015).

Tässä työssä esitetyt menetelmät ovat hyvin joustavia, ja niillä tehdyt kyselyt ovat helposti muokattavissa. Vaikeimmin toteutettavat, geometriset testit ovat tästä syystä johdettu standardeihin ja spatiaalisiin relaatioihin perustuen. Näiden menetelmien avulla pystytään toteuttamaan käytännössä mitä tahansa laatutestejä, ja miksei muitakin vektoriaineistoa hyödyntäviä paikkatietoanalyysyjä. Avoimeen lähdekoodiin perustuvat työkalut ovat helposti mukautettavissa, joten niitä voidaan soveltaa moneen tarpeeseen.

Työssä esitettyjen menetelmien avulla voidaan toteuttaa esimerkiksi ohjelmointikirjasto tai graafinen laatutestaustyökalu. Java-pohjaisiin toteutuksiin PostGIS tarjoaa valmiit ohjelmointikirjastot. PostGIS-geometriakohteita saadaan PostgreSQL-tietokannasta suoraan tekstimuodossa tai JDBC-laajennuksen oliona.

ELF-laadunarviointityökalun kehitys voisi yksinkertaisimmillaan olla melko helppoa, sillä PostGIS toteuttaa kaikki laadunarviointiin tarvittavat tekniset asiat. Työkalu voisi esimerkiksi lukea SQL-tiedoston ja suorittaa tietokannalle kyselyn käyttäjän antamalla aineistoparametreilla. Tietokannan yhteysasetukset ja aineistoparametrit voidaan tallentaa tiedostoksi myöhempää käyttöä varten. Tulokset voitaisiin palauttaa jossain sopivassa muodossa, esimerkiksi kaikki virheet sisältävänä Shapefilenä. Testit olisivat näin helposti muokattavissa SQL-tiedostoina, ja niitä voitaisiin myös helposti kirjoittaa aineistoteemoja ja eri tarkkuustasoja kattaviksi kokonaisuuksiksi. Laaturaportin muodostus olisi myös mahdollista tässä yhteydessä. Mitä enemmän käyttäjälle annetaan vapauksia, sitä monimutkaisemmaksi tulee työkalun kehitys. Kaavio yksinkertaisen laatutyökalun suhteista on kuvassa 22.



Kuva 22. Yksinkertainen kaavio yksinkertaisesta laatutyökalusta.

Automaattista laadunarviointia voitaisiin hyödyntää monessa paikkatietojen elinkaaren vaiheessa aina tuotannosta ylläpitoon. Tällä hetkellä menetelmää voidaan hyödyntää aina, kun saatavilla on vektoriaineistoa ja määrittelyt, joita sen tulee noudattaa. Vaikka tässä työssä laadunarviointi kohdistuukin koko tietoaaineistoon, voidaan arviointi kohdentaa myös pienempiin kokonaisuuksiin. Tuotantovaiheessa menetelmän avulla voitaisiin säännöllisesti tarkastella tehdyn työn tuloksia virheiden karsimiseksi ja tuotantomenetelmien tehostamiseksi. Aineistojen osia päivitettäessä voitaisiin tarkastella päivityksen kohteena olevien osa-alueiden tai kohteiden ja niiden ympäristöjen eheyttä ennen niiden hyväksymistä datasettiin. Tulevaisuuden visiossa menetelmää voisi ehkä hyödyntää reaaliaikaisesti tuotantoprosessien sisällä estämään virheiden syntyä globaaleihin laatusääntöihin verraten. Tämän työn konseptilla on potentiaalia siirtää paikkatiedon laadunhallinnan käytäntöjä uudelle aikakaudelle.

Lähdeluettelo

1Spatial. 2015. Products & Services. [Online]. [Viitattu 10.11.2015]. Saatavissa: <http://1spatial.com/products-services>.

2007/2/EY. 2007. Euroopan parlamentin ja neuvoston direktiivi 2007/2/EY Euroopan yhteisön paikkatietoinfrastruktuurin (INSPIRE) perustamisesta. Annettu 14.3.2007. Saatavissa: http://eur-lex.europa.eu/legal-content/FI/TXT/?uri=uriserv:OJ.L_.2007.108.01.0001.01.FIN.

922/2014. 2014. Valtioneuvoston asetus paikkatietoinfrastruktuurista annetun valtioneuvoston asetuksen 1 ja 5 §:n muuttamisesta. Maa- ja metsätalousministeriö. Saatavissa: <http://www.finlex.fi/fi/laki/alkup/2014/20140922>.

ArcGIS Resources. 2015. Data Reviewer. Esri. [Online]. [Viitattu 5.11.2015]. Saatavissa: <http://resources.arcgis.com/en/communities/data-reviewer/>.

Arroyo Ohori, K., Ledoux, H., & Meijers, M. 2012. Validation and automatic repair of planar partitions using a constrained triangulation. Photogrammetrie, Fernerkundung, Geoinformation (PFG). Vol. 5. S. 613-630. DOI: 10.1127/1432-8364/2012/0143. Saatavissa: https://3d.bk.tudelft.nl/ken/files/12_pfg.pdf.

Barça, J. 2015. How to detect gaps in a MultiPolygon table with Postgis? StackExchange communities: Geographic Information Systems. [Online]. 02.05.2015. [Viitattu 21.8.2015]. Saatavissa: <http://gis.stackexchange.com/questions/133134/how-to-detect-gaps-in-a-multipolygon-table-with-postgis>.

Beare, M. Henriksson, R., Jakobsson, A., Marttinen, J., Onstein, E., Tsoulos, L., Williams, F., Mäkelä, J., De Meulenaer, L., Persson, I., & Kavadas, I. 2010. D 8.4 ESDIN Quality Final Report - Part A. ESDIN consortium. Saatavissa: http://www.esdin.eu/sites/esdin.eu/files/D8-4_ESDIN_Quality_Final_Report.pdf.

Clementini, E. & Di Felice, P. 1995. A Comparison of Methods for Representing Topological Relationships. Information Sciences-Applications, Vol. 3. S. 149-178. Saatavissa: www.didattica.univaq.it/moodle/file.php/1508/InfoSci95.pdf.

Clementini, E., Sharma, J., & Egenhofer, M. J. 1994. Modeling Topological Spatial Relations: Strategies for Query Processing. Computers and Graphics. Vol 18:6. S. 815-822. Saatavissa: <http://www.spatial.maine.edu/~max/9intOptimization.pdf>.

Clementini, E., Di Felice, P., & van Oosterom, P., 1993. A Small Set of Formal Topological Relationships Suitable for End-User Interaction. Advances in Spatial Databases. S. 277-295. Springer Berlin Heidelberg. Saatavissa: <http://delab.csd.auth.gr/~alex/sdb/artSSD93.pdf>.

Delattre, N. 2015. D 2.3.1: Organizational Framework. The European Location Framework. Saatavissa: http://elfproject.eu/sites/default/files/D2.3.1_ELF_OrganisationalFramework.pdf.

Devillers, R. & Jeansoulin, R. 2006. Fundamentals of spatial data quality. ISTE Publishing Company. ISBN: 9781905209569 (painettu). ISBN: 9780470612156 (sähköinen). DOI: 10.1002/9780470612156.

Egenhofer, M. J., Sharma, J., & Mark, D. M. 1993. A critical comparison of the 4-intersection and 9-intersection models for spatial relations: formal analysis. Autocarto-conference 11. ASPRS American Society for Photogrammetry and Remote Sensing. 12 s. Saatavissa: <http://www.spatial.maine.edu/~max/4Vs9.pdf>.

ELF Feature Catalog. 2014. European Location Framework. [Online]. [Viitattu 21.10.2015]. Saatavissa: <http://elfproject.eu/documentation/specification/elf-data-specifications>.

FSF. 1991. GNU GENERAL PUBLIC LICENSE. Free Software Foundation, Inc. [Online]. [Viitattu 22.7.2015]. Saatavissa: <https://www.gnu.org/licenses/gpl-2.0.html>.

Garcia-Molina, H., Ullman, J. D., & Widom, J. (2009). Database systems: The complete book. 2nd ed. Upper Saddle River, New Jersey: Pearson Prentice Hall. 1203 s. ISBN: 978-0131873254.

GNU Wget. 2015. Introduction to GNU Wget. [Online]. [Viitattu 27.10.2015]. Saatavissa: <https://www.gnu.org/software/wget/>.

Hellerstein, J. M., Naughton, J. F., & Pfeffer, A. 1995. Generalized Search Trees for Database Systems. Proc. 21st Int'l Conf. on Very Large Data Bases, Zürich. S. 562-573. Saatavissa: <http://db.cs.berkeley.edu/papers/vldb95-gist.pdf>.

Häkli, P., Puupponen, J., Koivula, H., & Poutanen, M. 2009. Suomen geodeettiset koordinaatit ja niiden väliset muunnokset. Geodeettinen Laitos. Tiedote 30. 121 s. ISBN 978-951-711-274-1. ISSN 0787-9172. Saatavissa: <http://www.fgi.fi/fgi/sites/default/files/publications/gltiedote/GLtiedote30.pdf>.

INSPIRE. 2009. D2.8.I.1 INSPIRE Specification on Coordinate Reference Systems – Guidelines. Saatavissa: http://inspire.ec.europa.eu/documents/Data_Specifications/INSPIRE_Specification_CRS_v3.0.pdf.

INSPIRE. 2008. Deliverable 2.3: Definition of Annex Themes and Scope. Saatavissa: http://inspire.ec.europa.eu/reports/ImplementingRules/DataSpecifications/D2.3_Definition_of_Annex_Themes_and_scope_v3.0.pdf

ISO 9000. 2005. Quality management systems – Fundamentals and vocabulary. ISO/TC 176/SC 1. 52 s.

ISO 19157. 2013. Geographic information – Data Quality. ISO/TC 211. 146 s.

ISO/IEC 13249-3. 2011. Information technology – Database languages – SQL multimedia and application packages – Part 3: Spatial. ISO/IEC JTC 1/SC 32. 742 s.

Jakobsson, A. 2012. European Location Framework White Paper v1.0. [Viitattu 21.7.2015]. Saatavissa: <http://www.elfproject.eu/sites/default/files/ELF%20White%20Paper.pdf>.

JHS 154. 2012. ETRS89-järjestelmään liittyvät karttaprojektiot, tasokoordinaatistot ja karttalehtijako. JUHTA - Julkisen hallinnon tietohallinnon neuvottelukunta. 15 s. Saatavissa: <http://www.jhs-suositukset.fi/suomi/jhs154>.

JHS 158. 2010. Paikkatiedon metatiedot. JUHTA - Julkisen hallinnon tietohallinnon neuvottelukunta. 17 s. Saatavissa: <http://www.jhs-suositukset.fi/suomi/jhs158>.

JHS 160. 2006. Paikkatiedon laadunhallinta. JUHTA - Julkisen hallinnon tietohallinnon neuvottelukunta. 22 s. Saatavissa: <http://www.jhs-suositukset.fi/suomi/jhs160>.

JHS 169. 2012. Avoimen lähdekoodin ohjelmien käyttö julkisessa hallinnossa. JUHTA - Julkisen hallinnon tietohallinnon neuvottelukunta. 34 s. Saatavissa: <http://www.jhs-suositukset.fi/suomi/jhs169>.

Kania, J. 2013. Polygon to fill gap. StackExchange communities: Geographic Information Systems. [Online]. 14.5.2013. [Viitattu 21.08.2015]. Saatavissa: <http://gis.stackexchange.com/questions/60655/polygon-to-fill-gap/60672#60672>.

Ledoux, H., & Meijers, M. 2010. Validation of planar partitions using constrained triangulations. Proceedings ISPRS Commission II Mid-Term Symposium. Theory, Data Handling and Modelling in GeoSpatial Information Science, Hong Kong. IAPRS, XXXVIII. International Society of Photogrammetry and Remote Sensing (ISPRS). Saatavissa: http://repository.tudelft.nl/assets/uuid:6aac87a7-4e94-4b10-a8b4-f997529ff841/Ledoux_2010.pdf.

Obe, R., & Hsu, L. 2011. PostGIS in Action. Manning Publications Co. ISBN 978-1-935182-26-9.

Odoo.com forum. 2015. Reason why Odoo being slow when there is huge data inside the database. [Online]. 7.1.2015. [Viitattu 15.10.2015]. Saatavissa: <https://www.odoo.com/forum/help-1/question/reason-why-odoo-being-slow-when-there-is-huge-data-inside-the-database-87498>.

OGC 06-103r4. 2011. OpenGIS Implementation Standard for Geographic information - Simple feature access - Part 1: Common architecture. Version 1.2.1. 93 s. Saatavissa: http://portal.opengeospatial.org/files/?artifact_id=25355.

OGC 06-104r4. 2010. OpenGIS Implementation Standard for Geographic information - Simple feature access - Part 2: SQL option. Version 1.2.1. Open Geospatial Consortium. 111 s. Saatavissa: http://portal.opengeospatial.org/files/?artifact_id=25354.

OGC 09-025r1. 2010. OpenGIS Web Feature Service 2.0 Interface Standard. Version 2.0.0. Open Geospatial Consortium. 239 s. Saatavissa: http://portal.opengeospatial.org/files/?artifact_id=39967.

OpenGeo Suite 4.7. 2015. Spatial indexes. [Online]. [Viitattu 27.10.2015]. Saatavissa: <http://suite.opengeo.org/opengeo-docs/dataadmin/pgBasics/indexes.html>.

OSGeo. 2015. GeoTools. [Online]. [Viitattu 10.11.2015]. Saatavissa: <http://www.osgeo.org/geotools>.

Overton, D. 2011. D 1.10.3 ESDIN Public Final Project Report. ESDIN consortium. Saatavissa: http://www.esdin.eu/sites/esdin.eu/files/ESDIN%20D%201%2010%203%20Public%20Final%20Project%20Report%20v1.0_0.pdf.

PostGIS. 2015a. Binary installers. [Online]. [Viitattu 10.8.2015]. Saatavissa: <http://postgis.net/install>.

PostGIS. 2015b. Using PostGIS: Data Management and Queries. [Online]. [Viitattu 11.8.2015]. Saatavissa: http://postgis.net/docs/using_postgis_dbmanagement.html.

PostgreSQL. 2015. PostgreSQL 9.4.4 Documentation. The PostgreSQL Global Development Group. [Online]. [Viitattu 28.7.2015]. Saatavissa: <http://www.postgresql.org/docs/current/static/>.

PostgreSQL wiki. 2015. How does PostgreSQL use CPU resources? [Online]. [Viitattu 15.10.2015]. Saatavissa: https://wiki.postgresql.org/wiki/FAQ#How_does_PostgreSQL_use_CPU_resources.3F.

Ramsey, P. 2007. Introduction to PostGIS. Refrations Research Inc. Saatavissa: <http://2007.foss4g.org/workshops/W-04/PostGIS%20Workshop.doc>.

Ramsey, P. 2009. (Much) Faster Unions in PostGIS 1.4. Clever elephant, a perspicacious pachyderm. [Online]. 23.1.2009. [Viitattu 29.9.2015]. Saatavissa: <http://blog.cleverelephant.ca/2009/01/must-faster-unions-in-postgis-14.html>.

Ramsey, P. 2010. PostGIS - Tips for Power Users. FOSS4G 2010. Saatavissa: <http://2010.foss4g.org/presentations/3369.pdf>.

Refrations Research. 2015. PostGIS history. [Online]. [Viitattu 7.9.2015]. Saatavissa: <http://www.refrations.net/products/postgis/history/>.

Ross, R. G. 2009. RuleSpeak® Sentence Forms: Specifying Natural-Language Business Rules in English. Business Rule Solutions, LLC, the Business Rule Technique Company. Version 2.2.3. Saatavissa: www.rulespeak.com.

Ross, R. G. 2013. Tabulation of Lists in RuleSpeak®: Using “The Following” Clause. Business Rule Solutions, LLC, the Business Rule Technique Company. Saatavissa: www.rulespeak.com.

The PostGIS Development Group. 2015. PostGIS 2.1.9dev Manual. [Online]. [Viitattu 23.7.2015]. Saatavissa: <http://postgis.net/docs/manual-2.1/>.

TSK 45. 2014. Geoinformatiikan Sanasto. Sanastokeskus TSK ry. Saatavissa: <http://www.tsk.fi/tiedostot/pdf/GeoinformatiikanSanasto.pdf>. ISBN 978-952-9794-34-8. ISSN 1795-6323.

Vermeer, M. 2015. Käytännön geodesia. 133 s. Saatavissa: <http://users.aalto.fi/~mvermeer/kayt.pdf>.

Vivid Solutions. 2015a. JTS Javadoc: Geometry. [Online]. [Viitattu 6.10.2015]. Saatavissa: <http://www.vividsolutions.com/jts/javadoc/com/vividsolutions/jts/geom/Geometry.html>.

Vivid Solutions 2015b. JTS Topology Suite. [Online]. [Viitattu 10.11.2015]. Saatavissa: <http://www.vividsolutions.com/jts/JTSHome.htm>.

Werhahn, S., Jakobsson A., & Koistinen, K. 2015. Deliverable 2.3.6: Data Quality v0.6. The European Location Framework (ELF). Saatavissa: http://elfproject.eu/sites/default/files/D2.3.6_Quality.zip.

Liite 1. ELF-laatusäännöt ja laatuksissa käytetyt PostGIS-funktiot

RuleID	Quality Element	RuleSpeak	PostGIS functions	Comment
Geographical Names				
DEF01	-	A feature must be considered a <i>surface feature</i> if its <i>geometry</i> is of one of the following types *GM_Surface *GM_MultiSurface	-	Definition of a feature doesn't contain a test
DEF02	-	A feature must be considered a <i>line feature</i> if its <i>geometry</i> is of type GM_Curve	-	Definition of a feature doesn't contain a test
DEF03	-	A feature must be considered a <i>point feature</i> if its <i>geometry</i> is of type GM_Point	-	Definition of a feature doesn't contain a test
DEF04	-	A <i>NamedPlace</i> must be considered a <i>populated place</i> if its <i>type</i> is equal to <i>populatedPlace</i>	-	Definition of a feature doesn't contain a test
DEF05	-	A <i>populated place</i> must be considered a <i>point populated place</i> if its <i>geometry</i> is of type GM_Point	-	Definition of a feature doesn't contain a test
DEF06	-	A populated place must be considered a <i>surface populated place</i> if its <i>geometry</i> is of type GM_Surface	-	Definition of a feature doesn't contain a test
DEF07	-	A space between two or more surface features sharing a boundary must be considered a gap if it is surrounded by these features.	ST_Union ST_BuildArea ST_InteriorRingN ST_NumInteriorRings	Query done
DEF08	-	The thinness ratio of an area A with perimeter P must be computed as $4\pi \cdot \text{Area} / \text{Perimeter}^2$	ST_Area ST_Perimeter	Query done
DEF09	-	A gap must be considered a sliver if its thinness ration is smaller than or equal to 0,3. and its area is less than MaximumSliverArea	ST_Dump ST_Union ST_NumInteriorRings ST_MakePolygon ST_InteriorRingN ST_Area ST_Perimeter	Query done
GEN02.1 GEN02.2	Logical consistency format consistency	GEN2.1 A data set must be provided using the geographical reference system ETRS89-GRS80 (EPSG code 4258) if in the geographical scope of ETRS89 GEN2.2 A data set outside the geographical scope of ETRS89 must be provided using ITRS	ST_SRID	Query done

RuleID	Quality Element	RuleSpeak	PostGIS functions	Comment
GEN03	Logical consistency format consistency	A feature must conform to the definition of Simple Feature in OGC 06-103r4 if it is not in one of the following themes: *CP	ST_IsSimple	Query done
GEN04	Logical consistency conceptual consistency	The geometry of a feature must be all of the following: *not empty *minimum bounding rectangle / envelope not empty	ST_IsEmpty ST_Envelope	Query done
GEN05	completeness commission	The area of a surface feature must be greater than or equal to the MinimumAcceptedAreaSize .	ST_Area	Query done
GEN06	completeness commission	The length of a line feature must be greater than or equal to the MinimumLineLength	ST_Length	Query done
GEN07	Logical consistency topological consistency	The distance between two adjacent vertices of a feature must be greater than or equal to the MinimumVertexDistance .	ST_PointN ST_Npoints ST_MakeLine ST_Length ST_NumPoints	Query done for lines
GEN08.1 GEN08.2	Logical consistency topological consistency	GEN08.1 A node N must be considered 'Unnecessary' if all the following are true: * N intersects the end point of a line feature L * N intersects the end point of a line feature M * L and M are adjacent * The user-defined attributes of L are the same as those of M. GEN08.2 There must be no unnecessary nodes.	ST_EndPoint ST_Intersects ST_Touches ROW(a.1, a.2, a.3) = ROW(b.1, b.2, b.3)	Query done
GEN09	Logical consistency topological consistency	The minimum distance between each endpoint of a line L and another line M must be greater than OvershootUndershootLength if all the following are true: * L and M have the same type * No endpoint of L intersects M. (overshoots / undershoots)	ST_Intersects ST_DWithin ST_EndPoint	Query done
GEN11	Logical consistency topological consistency	A line feature must not overlap/intersect with another line feature of the same feature type if it is not in one of the following themes: *TN	ST_Overlaps	Query done

RuleID	Quality Element	RuleSpeak	PostGIS functions	Comment
GEN12	Logical consistency Conceptual consistency	A surface feature must not overlap/intersect with another surface feature of the same feature type if it is not in one of the following themes: *AU	ST_Overlaps	Query done
GEN13	Logical consistency topological consistency	There must be no slivers between areas of the same feature type.	ST_Union ST_BuildArea ST_InteriorRingN ST_NumInteriorRings	Query done
GEN14	Logical consistency topological consistency	A point feature must not overlap/intersect with another point feature of the same feature type	ST_Overlaps	Query done
GEN15	Logical consistency conceptual consistency	The value for endLifespanVersion must be one of the following: *later in time than the value for beginLifespanVersion B if B is not null *null	-	Query done Test compares column values
GEN16	Logical consistency format consistency	A feature with 2,5D geometry must be provided using the geographical reference system EVRS (EPSG:7.9.5. 5621) for its Z coordinate.	ST_SRID	No data
GEN50	-	The system must write the following statistics to the data quality report: * Number of features processed per feature type * Percentage of non-null values per attribute.	COUNT() ROUND() SUM()	Query done
GEN80	Logical consistency conceptual consistency	A feature must not cross the ELF InternationalBoundaries.	ST_Crosses	No data
GEN81	Logical consistency conceptual consistency	The intersection N of a feature with the ELF IB must be coincident with exactly one of the following if N is not null: * A Connecting Feature Point * A Connecting Feature Line.	ST_Intersection	No data

RuleID	Quality Element	RuleSpeak	PostGIS functions	Comment
Administrative Units				
ADM01.1 ADM01.2	Completeness Omission	<p>ADM01.1 A feature type must be considered 'Mandatory' if it is in the ExpectedFeaturesList.</p> <p>ADM01.2 At least one feature of each Mandatory feature type should be in the data set.</p>	-	Test attempts to find the first non-null value of a mandatory feature
ADM02	Logical consistency topological consistency	There must be no gaps between features that have the same nationalLevel value	(ADM09)	ADM09: There must be no gaps between (any) features
ADM03	Logical consistency topological consistency	<p>The attribute values of two area features A and B must be different for at least one user defined attribute if all the following are true:</p> <ul style="list-style-type: none"> * A and B are adjacent * A and B have the same feature type * A and B have the same nationalLevel value. 	ST_Intersects	Query done
ADM04	Logical consistency topological consistency	<p>An AdministrativeBoundary A must cover at least one edge of a feature F if all of the following are true:</p> <ul style="list-style-type: none"> * A and F have the same nationalLevel * F is one of the following types: <ul style="list-style-type: none"> ** AdministrativeUnit ** AdministrativeUnitArea. 	ST_Covers ST_ExteriorRing	No data
ADM06	Logical consistency topological consistency	Each edge of a AdministrativeUnit feature F must be covered by one AdministrativeBoundary if F and the AdministrativeBoundary feature have the same nationalLevel	ST_Covers ST_ExteriorRing	No data
ADM08	Logical consistency topological consistency	<p>Each end point of an AdministrativeBoundary A must intersect an end point of at least one AdministrativeBoundary B such that all the following are true:</p> <ul style="list-style-type: none"> * A and B are not the same feature * A and B have the same nationalLevel. 	ST_Intersects ST_EndPoint ST_StartPoint	Query done
ADM09	Logical consistency topological consistency	There must be no gaps between features	ST_Union ST_Dump ST_Difference ST_MakePolygon ST_ExteriorRing	Query done Overrides ADM02 GEN13 = LoD2->

RuleID	Quality Element	RuleSpeak	PostGIS functions	Comment
Buildings				
BUI01.1 BUI01.2	Completeness Omission	BUI1.1 A feature type must be considered 'Mandatory' if it is in the ExpectedFeaturesList BUI1.2.2 At least one feature of each Mandatory feature type should be in the data set.	-	Test attempts to find the first non-null occurrence of a mandatory feature
BUI02	Logical consistency format consistency	The attribute elevationValue of data type Elevation should be provided using EVRS (EPSG:7.9.5. 5621) for features within scope of EVRS	ST_SRID	No data, ETRS89+EVRF2007 EPSG:7423, EVRF2007 EPSG:5621
BUI03	Logical consistency format consistency	The value of the attribute lowReference of data type heightAboveGround must be one of the following: *generalGround *lowestGroundPoint *lowestFloorAboveGround *entrancePoint *highestGroundPoint *null	-	No data Test inspects attribute values
BUI04	Logical consistency domain consistency	The value of the attribute highReference of data type heightAboveGround must be one of the following: *generalRoofEdge *lowestRoofEdge *highestRoofEdge *generalEave *lowestEave *highestEave *generalRoof *top OfConstruction *highestPoint *null	-	No data Test inspects attribute values
BUI05	Logical consistency domain consistency	The value of the referenceGeometry attribute must be 'true' for exactly one geometric representation of a feature with data type BuildingGeometry2D	-	No data Test inspects attribute values

RuleID	Quality Element	RuleSpeak	PostGIS functions	Comment
BUI06.1 BUI06.2 BUI06.3	Logical consistency domain consistency	<p>BUI6.1: An attribute with one of the following names must be considered a 'BuildingEventDate':</p> <ul style="list-style-type: none"> * dateOfConstruction * dateOfDemolition * dateOfRenovation. <p>BUI6.2: The value of B.beginning for a BuildingEventDate B must be earlier than B.anyPoint.</p> <p>BUI6.3: The value of B.anyPoint for a BuildingEventDate B must be earlier than B.end.</p>	-	No data Test compares attribute values
BUI07	Logical consistency conceptual consistency	The value of NumberOfDwellings on Building must be equal to the sum of the values of number-OfDwellings of the BuildingParts composing the Building	-	No data Test compares attribute values
BUI08	Logical consistency conceptual consistency	The value of NumberOfBuildingUnits on Building must be equal to the sum of values of number-OfBuildingUnits of the BuildingParts composing the Building	-	No data Test compares attribute values
BUI09	Logical consistency domain consistency	The attribute horizontalGeometryEstimatedAccuracy of data type BuildingGeometry2D must be less than or equal to HorizontalGeometryEstimatedAccuracy	-	Query done Test compares attribute values
Cadastral Parcels				
CAD01	Logical consistency topological consistency	There must be no gaps between features	ST_Dump ST_Union ST_NumInteriorRings ST_MakePolygon ST_InteriorRingN	Query done
CAD02	Logical consistency topological consistency	Each end point of an Cadastral-Boundary must intersect an end point of at least one Cadastral-Boundary.	ST_Intersects ST_Relate	Query done

RuleID	Quality Element	RuleSpeak	PostGIS functions	Comment
Geographical Names				
GEO05	Logical consistency conceptual consistency	A PopulationRange must have all of the following attributes if it does not have a population: <ul style="list-style-type: none"> • lower • upper 	-	No data Test inspects attributes
GEO06	Logical consistency conceptual consistency	A PopulationRange must have a population if it does not have any of the following attributes: <ul style="list-style-type: none"> • lower • upper 	-	No data Test inspects attributes
GEO07.1 GEO07.2	Logical consistency conceptual consistency	GEO07.1 A PopulationRange must have a lower if it has an upper. GEO07.2 A PopulationRange must have an upper if it has a lower.	-	No data Test inspects attributes
GEO08	Logical consistency conceptual consistency	lower must be greater than 0 if upper is greater than 0	-	No data Test inspects attribute values
GEO09	Logical consistency conceptual consistency	lower must not be present if upper is not present	-	No data Test inspects attributes
GEO10	Logical consistency conceptual consistency	All of the following must be true if population is not present: <ul style="list-style-type: none"> • lower must be greater than or equal to 1 • upper must be greater than lower 	-	No data Test inspects attributes
GEO11	Logical consistency conceptual consistency	upper must be greater than lower if all of the following are true: <ul style="list-style-type: none"> • lower is not null • upper is not null 	-	No data Test inspects attributes

RuleID	Quality Element	RuleSpeak	PostGIS functions	Comment
GEO12.1 GEO12.2 GEO12.3	Logical consistency conceptual consistency	<p>GEO12.1 A language must be considered 'Not Latin' if it is one of the following: *BUL *GRE *SRP *UKR</p> <p>GEO12.2 A language must be considered 'Latin' if it is of value LAT</p> <p>GEO12.3 A NamedPlace N must have a GeographicalName G with G.language 'Latin' if N has a Geographical Name H with H.language 'Not Latin'</p>	-	Query done Test inspects attributes
GEO13	Completeness Omission	At least one feature of type NamedPlace should be in the data set.	-	Test attempts to find the first non-null occurrence of a mandatory feature
Hydrography				
HYD01	completeness commission	The area of an Island surface feature must be greater than or equal to the TargetAreaSize.	ST_Area	Query done
HYD02	completeness commission	The area of a Wetland surface feature must be greater than or equal to the TargetAreaSize.	ST_Area	Query done
HYD03	completeness commission	The area of a StandingWater surface feature must be greater than or equal to the TargetAreaSize.	ST_Area	Query done
HYD04.1 HYD04.2	Completeness Omission	<p>HYD04.1 A feature type must be considered 'Mandatory' if it is in the ExpectedFeaturesList</p> <p>HYD04.2 At least one feature of each Mandatory feature type should be in the data set.</p>	-	Test attempts to find the first non-null occurrence of a mandatory feature

RuleID	Quality Element	RuleSpeak	PostGIS functions	Comment
HYD05.1 HYD05.2 HYD05.3	Logical consistency topological consistency	<p>HYD05.1 A point feature with type in the WaterPointFeatureList must be considered a Water Point Feature.</p> <p>HYD05.2 A feature with type in the following list must be considered a Watercourse Feature:</p> <ul style="list-style-type: none"> * Watercourse * WatercourseLink * WatercourseSeparatedCrossing. <p>HYD05.3 A Water Point Feature must intersect at least one Watercourse Feature</p>	ST_Intersects	Query done
HYD06	Logical consistency conceptual consistency	<p>HYD06 The value of the boolean IsAverageWidthAtLeastMinimumWidth may be computed as follows:</p> <ol style="list-style-type: none"> 1. Create a negative buffer N of F of width equal to half of MinimumWidth. 2. Subtract N from F to produce an intermediate result R1. 3. <ol style="list-style-type: none"> a. If R1 consists of two or more disconnected surfaces, return "false". b. Otherwise, create a positive buffer P around R1 of width equal to (half of MinimumWidth) x 101%. 4. Add P to R1. 5. <ol style="list-style-type: none"> a. If the intersection of the final result R2 and F is equal to F, return "true". b. Otherwise, return "false" 	ST_Difference ST_Buffer ST_Union ST_GeometryType ST_Intersection	Performance issues
HYD07.1 HYD07.2 HYD07.3 HYD07.4	Logical consistency topological consistency	<p>HYD07.1 A feature with type in the following list must be considered a Water Barrier:</p> <ul style="list-style-type: none"> * Lock * DamOrWeir. <p>HYD07.2 A feature with type in the following list must be considered a Water Surface:</p> <ul style="list-style-type: none"> * Watercourse * StandingWater. <p>HYD07.3 A Water Barrier line feature must intersect the boundary of a Water Surface.</p> <p>HYD07.4 The intersection N of a Water Barrier line feature B and a Water Surface boundary must be equal to B if N is not null.</p>	ST_Intersects ST_Intersection ST_Equals	Query done

RuleID	Quality Element	RuleSpeak	PostGIS functions	Comment
HYD08	Logical consistency topological consistency	A Water Barrier point feature must intersect an endpoint of a Watercourse line feature.	ST_Intersects ST_EndPoint	Query done
HYD09	Logical consistency topological consistency	A Watercourse surface feature must contain at least one WatercourseLink if all of the following are true: *it has at least one ingoing watercourse *it has at least one outgoing watercourse	ST_EndPoint ST_StartPoint ST_Contains	No data
HYD10	Logical consistency topological consistency	A StandingWater surface feature must contain at least one WatercourseLink if all of the following are true: *it has at least one ingoing watercourse *it has at least one outgoing watercourse	ST_EndPoint ST_StartPoint ST_Contains	No data WatercourseLinks stop at waterbody borders
HYD11	Logical consistency topological consistency	A Wetland surface feature must contain at least one WatercourseLink if all of the following are true: *it has at least one ingoing watercourse *it has at least one outgoing watercourse	ST_EndPoint ST_StartPoint ST_Contains	No data
HYD12	Logical consistency topological consistency	There must be a node at each point of intersection between an intersecting pair of features from LineFeaturesList	ST_Intersection ST_Equals	No data
HYD13	Logical consistency topological consistency	A point feature from the PointFeaturesList must not be within a surface of one of the following feature types: *StandingWater *Watercourse *SeaArea *Sea	ST_Contains	Query done
HYD14	Logical consistency topological consistency	A Coastline feature must be covered by the boundary of one or more SeaArea features.	ST_Covers ST_ExteriorRing	No data

RuleID	Quality Element	RuleSpeak	PostGIS functions	Comment
HYD15	Logical consistency topological consistency	Exactly one of the following must be true for a LandWaterBoundary feature B: * B is closed * Each end point of B intersects the end point of another LandWater-Boundary feature.	ST_Intersects ST_EndPoint ST_StartPoint	No data
HYD16	Logical consistency topological consistency	Adjacent features of different feature types must not have gaps that are slivers.	ST_Union ST_BuildArea ST_InteriorRingN ST_NumInteriorRings	Query done
HYD17	Logical consistency topological consistency	A SeaArea feature must not overlap with a feature of any of the following types: *StandingWater *Watercourse *Wetland *Island	ST_Overlaps	Query done
HYD18	Logical consistency topological consistency	A StandingWater feature must not overlap with a feature of any of the following types: *Watercourse *Wetland *Island	ST_Overlaps	Query done
HYD19	Logical consistency topological consistency	A Watercourse feature must not overlap with a feature of any of the following types: *Wetland *Island	ST_Overlaps	No data
HYD20.1 HYD20.2	Logical consistency topological consistency	HYP20.1 A feature must be considered 'Foreshore' if it is one of the following types: *Shore *InterTidalArea HYP20.2 A Foreshore feature must overlap at least one of the following types: *SeaArea *Watercourse.	ST_Overlaps	No data
HYD21	Logical consistency topological consistency	A Foreshore feature must not overlap with an Island feature.	ST_Overlaps	No data
HYD22	Logical consistency topological consistency	A WatercourseLink feature must be within the union of one or more of the following surface features: *Watercourse *StandingWater	ST_Within ST_Union	Query done

RuleID	Quality Element	RuleSpeak	PostGIS functions	Comment
HYD23	Logical consistency conceptual consistency	The tidal attribute of a Watercourse feature W must be 'False' if any of the following is true: * W.origin is 'manMade' * W.level is 'suspendedOrElevated' * W.level is 'underground'.	-	No data Test inspects attribute values
HYD26.1 HYD26.2 HYD26.3	Logical consistency domain consistency	HYD26.1 The width.lower value of a Watercourse feature W must be considered the 'Lower Width' of W. HYD26.2 The width.upper value of a Watercourse feature W must be considered the 'Upper Width' of W. HYD26.3 The Lower Width of a Watercourse feature W must be less than the Upper Width of W.	-	No data Test compares attribute values
HYD27.1 HYD27.2	Logical consistency domain consistency	HYD27.1 A feature must be considered 'Fresh Water' if it is one of the following types: * Watercourse * StandingWater. HYD27.2 The tidal attribute of a Fresh Water feature F must be 'False' if F.persistence is one of the following values: * 'dry' * 'ephemeral' * 'intermittent'	-	No data Test inspects attribute values
HYD28	Logical consistency conceptual consistency	The level attribute of a Watercourse feature W must not be equal to 'suspendedOrElevated' if W.origin is equal to 'natural'.	-	No data Test inspects attribute values
HYD29	Logical consistency topological consistency	A Water Barrier line feature must intersect a Water surface feature. (See HYD07)	ST_Intersects	Query done
HYD31.1 HYD31.2	Logical consistency topological consistency	HYD31.1 A feature must be considered a MaritimShoreline if it is of one of the following types: * Shoreline * Coastline HYD31.2 Exactly one of the following must be true for a MaritimShoreline feature B: * B is closed * Each end point of B intersects the end point of another MaritimShoreline feature.	ST_IsClosed ST_Intersects	No data

RuleID	Quality Element	RuleSpeak	PostGIS functions	Comment
Transport Network				
TRA01.1 TRA01.2	Completeness Omission	<p>TRA01.1 A feature type must be considered 'Mandatory' if it is in the ExpectedFeaturesList:</p> <p>TRA01.2 At least one feature of each Mandatory feature type should be in the data set.</p>	-	Test attempts to find the first non-null occurrence of a mandatory feature
TRA02	Logical consistency topological consistency	Each AerodromeArea surface feature must contain exactly one AerodromeNode point feature	ST_Contains	Query done
TRA03	Logical consistency topological consistency	Each AerodromeNode feature must intersect the end point of at least one feature of one of the following types: *RoadLink *RailwayLink	ST_Intersects ST_EndPoint	No data
TRA04	Logical consistency topological consistency	A line feature of RunwayLine must lie within an AerodromeArea surface feature	ST_Within	No data
TRA05	Logical consistency topological consistency	At least one end point of a RoadLink must intersect an endpoint of at least one other RoadLink.	ST_Intersects ST_EndPoint ST_StartPoint	Query done
TRA06	Logical consistency topological consistency	At least one end point of a RailwayLink must intersect an endpoint of at least one other RailwayLink	ST_Intersects ST_EndPoint	Query done
TRA07	Logical consistency topological consistency	Each SchengenBorderPoint feature must intersect the end point of one or more features of the following types: *RoadLink *RailwayLink	ST_EndPoint ST_StartPoint ST_Intersects	No data (SchengenBorderPoint)
TRA08	Logical consistency topological consistency	Each end point of a FerryCrossing feature must intersect exactly one PortNode feature.	ST_Intersects ST_EndPoint ST_StartPoint	No data

RuleID	Quality Element	RuleSpeak	PostGIS functions	Comment
TRA09	Logical consistency topological consistency	Each PortNode feature must intersect the end point of at least one feature of the following types: *RoadLink *RailwayLink	ST_EndPoint ST_StartPoint ST_Intersects	No data
TRA10	Logical consistency topological consistency	A PortArea must contain exactly one PortNode	ST_Contains	No data
TRA11	Logical consistency topological consistency	Each InterchangePoint feature must intersect the end point of at least two RoadLink features	ST_Intersects ST_EndPoint ST_StartPoint	No data (InterchangePoint)
TRA12	Logical consistency topological consistency	Each RoadNode feature must intersect the end point of at least one RoadLink feature	ST_Intersects ST_EndPoint ST_StartPoint	No data
TRA13	Logical consistency topological consistency	Each RailwayStationNode feature must intersect the end point of at least one feature of each of the following types: *RoadLink *RailwayLink	ST_EndPoint ST_StartPoint ST_Intersects	Query done
TRA14	Logical consistency topological consistency	Each RailwayNode feature must intersect the end point of at least one RailwayLink feature	ST_Intersects ST_EndPoint ST_StartPoint	Query done
TRA15	Logical consistency topological consistency	Each RoadNode feature must intersect at least one RoadLink feature	ST_Intersects	No data
TRA16	Logical consistency topological consistency	Each RailwayStationNode feature must intersect at least one RailwayLink feature	ST_Intersects	Query done

RuleID	Quality Element	RuleSpeak	PostGIS functions	Comment
Protected Sites				
PRO02	Completeness Omission	At least one feature of Protected-Site should be in the data set.	-	Test attempts to find the first non-null occurrence of a mandatory feature
Land Cover				
LAN01	completeness commission	The area of a surface feature must be greater than or equal to the TargetAreaSize.	-	ST_Area is not needed because these features include pre-calculated areas
LAN02	completeness omission	At least one feature of LandCover-Unit should be in the data set.	-	Test attempts to find the first non-null occurrence of a mandatory feature
Cross-theme				
CRO01.1 CRO01.2 CRO01.3 CRO01.4	Logical consistency topological consistency	<p>CRO01.1 A buffer of width IBbuffer-Width around the IB must be considered the 'IB Buffer'.</p> <p>CRO01.2 A Watercourse W must be considered a 'Boundary Water Feature' if all the following are true: * At least one segment S of W is within the IB Buffer * The length of S is greater than MinBoundarySegmentLength.</p> <p>CRO01.3 An AdministrativeBoundary feature A must be considered '1st Order' if A.nationalLevel is equal to '1stOrder'.</p> <p>CRO01.4 A Boundary Water Feature must be covered by a 1st Order AdministrativeBoundary</p>	ST_Buffer ST_Intersects ST_MakeLine ST_PointN ST_NumPoints ST_Within ST_Length ST_Covers	No data
CRO02.1 CRO02.2	Logical consistency topological consistency	<p>CRO02.1 A NamedPlace feature S must be considered a SettlementPoint if it has S.type 'populated Place'</p> <p>CRO02.2 A surface feature must be considered a WaterBody if it is in the WaterBodyList.</p> <p>CRO02.3 A SettlementPoint feature must not be within a WaterBody feature</p>	ST_Within	No data

RuleID	Quality Element	RuleSpeak	PostGIS functions	Comment
CRO03	Logical consistency topological consistency	A LandCoverUnit feature must not overlap a WaterBody feature	ST_Overlaps	Query done
CRO04.1 CRO04.2 CRO04.3	Logical consistency topological consistency	CRO04.1 A feature must be considered a TransportFeature if it is in the TransportFeatureList CRO04.2 A WaterBody feature must not overlap a TransportFeature. CRO04.3	ST_Overlaps	Query done
CRO05	Logical consistency topological consistency	A SchengenBorderPoint must intersect a 1stOrder AdministrativeBoundary	ST_Intersects	No data
CRO07	Logical consistency topological consistency	A FerryCrossing feature must be within at least one WaterBody feature	ST_Union ST_Intersects ST_Within	Query done
CRO08.1 CRO08.2	Logical consistency topological consistency	CRO08.1 An AdministrativeUnitArea A must be considered an 'Inland Water Feature' if A.landCoverType is equal to 'inland-Water'. CRO08.2 An Inland Water Feature must be covered by a Standing-Water feature.	ST_Covers	No data
CRO09	Logical consistency topological consistency	An ELF InternationalBoundary must be covered by a 1st Order AdministrativeBoundary feature.	ST_Covers	No data
CRO10	Logical consistency topological consistency	A SettlementPoint feature must not be within any of the following types: *AerodromeArea *PortArea	ST_Within	No data
CRO11	Logical consistency topological consistency	The distance between a SettlementPoint feature and a RoadLink feature must be less than or equal to the MinimumSettlementDistance	ST_DWithin	No data

RuleID	Quality Element	RuleSpeak	PostGIS functions	Comment
CRO12.1	Logical consistency conceptual consistency	CRO12.1 A LandCoverUnit L must be considered a SettlementArea if L.LandcoverObservation is 'artificial Construction'		
CRO12.2		CRO12.2 A SettlementPoint P must be considered a SettlementAreaPoint if it is within a SettlementArea.		
CRO12.3		CRO12.3 The area of a SettlementArea feature must be greater than or equal to MinimumArea		
CRO12.4		CRO12.4 A SettlementAreaPoint P must have one of the following: * P.population is greater than or equal to ThresholdInhabitants * P.lower is greater than or equal to ThresholdInhabitants	ST_Area ST_Within	No data